

Complexity in Chemistry, Biology, and Ecology

Edited by

Danail Bonchev

Virginia Commonwealth University
Richmond, Virginia

and

Dennis H. Rouvray

University of Georgia
Athens, Georgia



Springer

VISIT...

LANZAROTE
Caliente.COM

Library of Congress Control Number: 2005925502

ISBN-10: 0-387-23264-8

eISBN: 0-387-25871-X

ISBN-13: 978-0387-23264-5

Printed on acid-free paper

© 2005 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (TB/EB)

9 8 7 6 5 4 3 2 1

springeronline.com

CONTRIBUTORS

Alexandru T. Balaban, Texas A & M University at Galveston, Galveston, Texas

Danail Bonchev, Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia

Gregory A. Buck, Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia

Pau Fernández, ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra (GRIB), Barcelona, Spain

Gabor Forgacs, University of Missouri, Columbia, Missouri

Xiaofeng Guo, Department of Mathematics, Xiamen University, P. R. China

Lemont B. Kier, Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia

Donald C. Mikulecky, Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia

Stuart A. Newman, New York Medical College, Valhalla, New York

Dejan Plavšić, Institute Rudjer Bošković, Zagreb, Croatia

Milan Randić, National Institute of Chemistry, Ljubljana, Slovenia

Ricard V. Solé, ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Barcelona, Spain

Robert E. Ulanowicz, University of Maryland, Center for Environmental Science, Chesapeake Biological Laboratory

Tarynn M. Witten, Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia

PREFACE

As we were at pains to point out in the companion volume to this monograph, entitled *Complexity in Chemistry: Introduction and Fundamentals*, complexity is to be encountered just about everywhere. All that is needed for us to see it is a suitably trained eye and it then appears almost magically in all manner of guises. Because of its ubiquity, complexity has been and currently still is being defined in a number of different ways. Some of these definitions have led us to major and powerful new insights. Thus, even in the present monograph, the important distinction is drawn between the interpretations of the concepts of complexity and complication and this is shown to have a significant bearing on how systems are modeled. Having said this, however, we should not fail to mention that the broad consensus that now gained acceptance is that all of the definitions of complexity are in the last analysis to be understood in essentially intuitive terms. Such definitions will therefore always have a certain degree of fuzziness associated with them. But this latter desideratum should in no way be viewed as diminishing the great usefulness of the concept in any of the many scientific disciplines to which it can be applied. In the chapters that are included in this monograph the fact that differing concepts of complexity can be utilized in a variety of disciplines is made explicit. The specific disciplines that we embrace herein are chemistry, biochemistry, biology, and ecology.

Chapter 1, “On the Complexity of Fullerenes and Nanotubes,” is written by an international team of scientists led by Milan Randić. While devoted to specific chemical applications of complexity theory, and dealing

with hot topics in contemporary chemical technology, the chapter complements contributions to the quantification of complexity made in the preceding volume *Complexity in Chemistry*. Most approaches to the notion of complexity in molecules and molecular graphs have been based on an evaluation of selected graph invariants, calculated for the graph itself or, most recently, for all of its subgraphs. This chapter focuses more on the influence that symmetry elements have on complexity of the objects considered. This is a controversial theme, with opposing opinions in the literature, because of the prevailing view on symmetry as a simplifying factor. The authors offer an improvement of symmetry-based complexity measures by accounting for the cardinality of the sets of equivalent elements. In addition, the concept of presenting the complexity measure as a complexity vector or sequence, originally developed for subgraph-based complexity measures, is now realized for distance-based sequences. The latter contain the average count of the number of nearest neighbors at various distances, and in the case of the fullerenes also the average distance between the twelve pentagonal faces of the fullerenes. The review ends with a discussion of the complexity of nanotubes, for which the main role appears to be played by the twist and counter-twist parameters that determine the nanotube helicity and diameter. As in the case of the fullerenes, the authors conclude that no single parameter seems to be sufficient to characterize nanotube complexity.

In Chapter 2, Newman and Forgacs focus on the physicochemical aspects of complexity in developmental and evolutionary biology. The major emphasis in development studies has traditionally been on the hierarchical regulatory relationships among genes, while the variation of genes has played a corresponding role in evolutionary research. Recently, however, investigators have focused on the roles played by the physical and dynamical properties of cells and tissues in producing biological characteristics during ontogeny and phylogeny. The interactions among gene products, metabolites, ions, etc., reaction-diffusion coupling, and opportunities for molecular diffusion over macroscopic distances, lead to self-organizing multistable, oscillatory, and pattern forming dynamics. These system properties, not specified in any genetic program, can account for most of the features of animal body structures, including cell differentiation, tissue multilayering, segmentation, and left-right asymmetry. The authors point out that these chemical-dynamic properties are *generic* and are common to living and nonliving systems. As such, they have played a major role in the evolution

of ancestral multicellular organisms, more than for modern organisms with their hierarchical genetic control. The morphologies originally generated by physicochemical dynamics probably provided morphological templates for genetic evolution that stabilized and reinforced (rather than innovated) multicellular body plans and organ forms. The hierarchical genetic control of development seen in modern organisms can thus be considered as an outcome of this evolutionary interplay between genetic change and generic physicochemical processes.

The chapter “The Circle That Never Ends: Can Complexity Be Made Simple?” by Mikulecky begins with the premise that all real systems are complex and that there can be no single approach to such systems. Mikulecky presents an introduction to a relatively new approach called “relational systems theory” to which he has made valuable contributions along with such pioneers as Katchalsky, Peusner, and Rosen. As an essential part of the efforts being made to create a new, non-Newtonian paradigm of science, this theory accounts for the irreducibility of certain context dependent functional components in the system, components that would disappear when the context defining them is destroyed by reductionist techniques. The influence of reductionism on the methods of science is described by Mikulecky in some detail in an effort to provide reasons for moving beyond the restrictions to systems imposed on us by this traditional approach. The relational approach clearly identifies the nature of the *functional* components as that entity which makes a complex whole more than the sum of its parts. The relational model is expressed in terms of processes rather than its physical parts. There is no way to uniquely identify a functional component by the way it relates to the physical parts of the system even though a relationship has to exist. Relational models that are context dependent and self-referential are considered inherently incapable of being reduced to the algorithmic procedures that make mechanistic systems so adaptable to computer simulation and other computational techniques. This non-computability, which makes it impossible to explain fully a system proceeding from a single model, is thus regarded as a key characteristic of complex reality.

The study of the organization extant in mechanisms holds out the promise of bridging the gap between the mechanistic approach characteristic of reductionism and the new relational approach. For that reason, Mikulecky makes use of Network Thermodynamics to illustrate the relation between organization and the material parts in mechanisms as a way of showing

that the material parts alone are insufficient to give a system description even for mechanisms. Network Thermodynamics combines topology with analytical mathematics to model large complicated systems in a way that demonstrates the role of organization in dynamic models. Versions of Network Thermodynamics have been developed during the last 30 years by Oster, Perelson and Katchalsky, Peusner, and Mikulecky. Summarizing these results, this review focuses on a generalized formalism of electrical circuit theory, which is shown to be applicable to all networks representing physical systems. The reader can thus gain new and fruitful insights into the possibilities for modeling complex dynamic networks. One can only agree with Mikulecky's conclusion that "The challenge arising from acknowledging the complexity of the real world is to try to maintain scientific rigor without stripping the investigative process of the tools needed to deal with context dependence and self-reference. In that way, the circle never ends."

Complex biochemical systems are difficult to study in their entirety due to the overwhelming number of constituents they have. However, in focusing on the interactions between the constituents one arrives at the underlying networks, the best representation of the integrated whole. This is the starting point in Chapter 4, which is devoted to networks (graphs) as models of large-scale biochemical organization. With many examples, from the contact network in a folded protein to protein-protein interaction networks to gene regulatory networks, it is convincingly demonstrated how some of the properties of complex systems can be approached through the study of the properties of their corresponding networks. One of the very few reviews on the topological properties of complex networks, this article begins by introducing the necessary notions of graph theory. Following the chronology of development of complex network theory, the authors define first the properties of random graphs. Vertex degree distribution is analyzed in detail, including clustering as a basic property of complex graphs. The transition from random graphs to apparently nonrandom, real-life networks is elegantly presented beginning with the random "long-distance shortcuts", and then going on to Strogatz and Watts' "small world" theory. The folded protein structure is the first example of a complex biochemical network the authors describe. The linear chain of residues is folded and this final shape involves contact, long-range interactions among the aminoacid residues. The mandatory small-world, connectivity and clustering analysis is followed by that of hierarchical clustering, which reveals the hidden modular

structure of the network. Protein-protein networks follow with examples of such network fragments in humans and *Saccharomyces cerevisiae*. Here, the latest techniques of network assortativeness and correlation profiles are introduced. The proteome emergence and evolution is modeled proceeding from an analogy with the gene duplication mechanism in genetic evolution. The last section covers the gene regulatory networks that determine the functioning of the living cell. These networks are analyzed as Kauffman's Boolean networks and possess a number of dynamic properties such as the emergence of high-order, robustness, and phase transitions. Three distinct phases are shown to exist in such networks: ordered, chaotic, and critical and the conditions for their existence are mathematically formulated. The formalism is illustrated with the regulatory networks of *E. coli* and *S. cerevisiae*. Using networks as a theoretical framework for complex biochemical systems is relevant for a number of reasons, and especially for providing well-defined quantitative properties to be reproduced by dynamical models of network evolution, thereby serving as a blueprint for the laws of organization of living matter.

Complexity of molecules and the criteria for quantitative complexity measures have been discussed in detail in the preceding volume. By this reason, Chapter 5, "How to Measure Complexity?", only briefly reviews the history of the topic. The emphasis is put on the latest development that identified nonrandom dynamic networks as a universal language to describe complexity of systems as diverse as the discrete space-time, molecules, living matter, ecosystems, social systems, financial markets, and World Wide Web. This revolutionary idea, advanced by Barabási in 1999, had a profound impact on life sciences and first of all on mathematical biology, which is more and more dominated by systems biology and bioinformatics. This has also changed the approaches used to quantify complexity in biology and ecology. In addition to information theory, traditionally used to define complexity as *compositional* diversity, graph theory emerged as a universal tool for assessing *structural* or *topological* complexity of any dynamic evolutionary system. The authors proceed with analysis of the role of symmetry as simplifying factor, diminishing systems complexity. They warn against the blind use of information theory for complexity estimates, and advocate the use of the information on the vertex degree distribution as one of the very few satisfactory information measures of topological complexity. A general scheme is presented for quantifying complexity as global, average, and normalized complexity indices. Three graph theoretical

measures of network complexity are recommended: the subgraph count, the overall connectivity index, and the walk count. All three are presented both as a global index, and as an ordered sequence of terms related to subgraphs, and respectively walks, from the smallest to the largest size. This makes it possible to assess networks complexity by the first several terms of each of these indices, avoiding thus the combinatorial explosion for large-scale networks. Two complexity measures combining graph adjacency and graph distance (graph radius) are also presented for the first time. These indices unite the intuitive ideas of structural complexity resulting from high connectivity and small vertex separation (the “small world” concept). Complexity of directed networks is analyzed by introducing a measure for vertex accessibility, which enables producing realistic total distance and connectedness estimates of these networks. The mathematical formalism introduced throughout the chapter is illustrated in detail by examples of protein-protein networks and food webs.

Chapter 6, “Cellular Automata Models of Complex Biochemical Systems,” begins by introducing the basic notions of systems, defining the states of the system, the system observables and their interactions. The basic features of modeling and simulation are discussed. Focusing on modeling in chemistry and molecular biology, two powerful methods for chemical investigations are discussed and compared: molecular dynamic and Monte Carlo simulations. Both approaches have great strengths and often lead to quite similar results for the properties of the systems studied. However, these methods depend on rather elaborate models of the molecular interactions. As a result, both methods are highly demanding computationally, and research-level calculations are normally run on supercomputers, clusters, or other large systems. Before introducing the alternative approach of *cellular automata* that greatly simplifies the view of the molecular system, and significantly reduces the computational demand, the authors first address the general principles of complexity. The notion of a complex system is distinguished from that of a complicated system, which is no more than the sum of its parts. In contrast, in complex systems “the whole is greater than the sum of the parts.” What is more these systems possess the properties of *emergence*, *adaptation*, and *self-organization*. The authors discuss these properties of complex systems along with the existing hierarchy of complexity focusing on the structure and interconnectedness of the “layers” of complexity. The contributions of the pioneers of complexity theory are elucidated and illustrated by numerous examples.

This chapter continues by introducing the basic notions and principles of cellular automata. Cellular automata consist of a discrete lattice of cells and evolve in discrete time steps. Each site takes on a finite number of possible values. The value of each site evolves according to the same simple, heuristic rules, which depend only on a local neighborhood of sites around it. The results of a CA model are new sets of states of the constituents called the configuration of the system. This configuration arises from many changes and encounters among the constituents of the CA, which may occur over a very long period of “time” in the model. The last part of the chapter presents abundant examples of applications of cellular automata to chemistry and biology, taken from Kier’s laboratory. The early work was directed toward the study of water and solution phenomena. Studies include cellular automata models of water as a solvent, dissolution of a solute, solution phenomena, the hydrophobic effect, oil and water de-mixing, solute partitioning between two immiscible solvents, micelle formation, diffusion in water, membrane permeability, acid dissociation, and dynamic percolation. Later work involved cellular automata models of molecular bond interactions, diffusion in water, including drug molecule diffusion and the hydrophobic effect, as well as generalizations of Kier’s chreode theory of diffusion in water and his theory of volatile anesthetic action. The chapter ends with a demonstration of the full potential of this powerful technique for application to complex biochemical pathways.

Over the last three centuries, Newtonian dynamics has strongly guided how we look at nature. Newtonian systems are explained in terms of mechanistic or material causes only. They are deterministic, reversible in time, decomposable into their parts and composable again. Physical laws are assumed to apply everywhere, at all times and over all scales. In Chapter 7, Ulanowicz shows that none of these notions applies to ecodynamics. His conclusion stems from the fact that constraints in living systems are not rigidly mechanical in nature, owing mainly to the cyclical relationships among some of them. Living systems are not fully constrained, i.e., they retain sufficient flexibility to adapt to changing circumstances. Flexibility is probably easier to discern in ecosystems than in organisms where the constraints are more prevalent and rigid. Autocatalysis plays an essential role in the emergence of non-mechanical behavior in all living systems. Whenever two or more autocatalytic loops arise from the same pool of resources, autocatalysis induces competition and symmetry-breaking. Ulanowicz asserts that the full ontogenetic mapping from genome to phenome is very

likely a chimera. What is really needed is try to discover the principles of self-organization. The author shows that the imbalances in material and energy demanded by physics usually equilibrate at rates that are much faster than changes occurring in the constraints that actually determine the pattern of system exchanges. Thus, the pathway to achieving a mathematical description of ecodynamics lies in the quantification of internal constraints. The control of ecodynamics appears to be relational in nature—how much any change in one constraint affects others with which it is linked. While it is impossible to treat explicitly all the constraints hidden within an ecosystem, their overall effects on the network of trophic interactions can nevertheless be quantified in a manner similar to that used in thermodynamics. For this purpose, Ulanowicz uses information theory arguments to introduce the concept of *ascendance* as a quantitative measure of how efficiently and coherently the system processes its medium. A phenomenological principle is proposed that “in the absence of major perturbations, ecosystems have a propensity to increase in ascendancy.” The ensuing description of ecodynamics, however, does not accord with the normal assumptions in Newtonian metaphysics, and Ulanowicz offers a reformulated ecological metaphysics in its place.

In conclusion, we would like to express the hope that the chapters contained within our monograph will not only make for some stimulating reading but that they will also afford our readers with valuable resource material for future research endeavors. We take this opportunity to thank all of our contributors for their valuable contributions to the fascinating subject of complexity. We can only hope that our readers will derive much pleasure in perusing the exciting offerings herein.

DANAIL BONCHEV
DENNIS H. ROUVRAY

CONTENTS

1. ON THE COMPLEXITY OF FULLERENES AND NANOTUBES	1
<i>Milan Randić, Xiaofeng Guo, Dejan Plavšić and Alexandru T. Balaban</i>	
1. Introduction	1
2. On the Complexity of the Complexity Concept	3
3. Complexity and Branching	4
4. Complexity of Smaller Molecules	7
5. Augmented Valence as a Complexity Index	16
6. Complexity of Smaller Fullerenes	20
7. Comparison of Local Atomic Environments	25
8. The Role of Symmetry	29
9. Concluding Remarks on the Complexity of Fullerenes	34
10. On the Complexity of Carbon Nanotubes	36
10.1. Introductory remarks	36
10.2. Helicity of nanotubes	38
Acknowledgement	43
References	44
2. COMPLEXITY AND SELF-ORGANIZATION IN BIOLOGICAL DEVELOPMENT AND EVOLUTION	49
<i>Stuart A. Newman and Gabor Forgacs</i>	
1. Introduction: Complex Chemical Systems in Biological Development and Evolution	49
2. Dynamic, Multistability and Cell Differentiation	51
2.1. Cell states and dynamics	53

2.2.	Epigenetic multistability: the Keller autoregulatory transcription factor network model	55
2.3.	Dependence of differentiation on cell-cell interaction: the Kaneko-Yomo “isologous Ddiversification” model	59
3.	Biochemical Oscillations and Segmentation	65
3.1.	Oscillatory dynamic oscillations and somitogenesis . .	65
3.2.	The Lewis model of the somitogenesis oscillator	66
4.	Reaction-Diffusion Mechanisms and Embryonic Pattern Formation	70
4.1.	Reaction-diffusion systems	71
4.2.	Axis formation and left-right asymmetry	71
4.3.	Meinhardt’s models for axis formation and symmetry breaking	72
5.	Evolution of Developmental Mechanisms	76
5.1.	Segmentation in insects	77
5.2.	Chemical dynamics and the evolution of insect segmentation	80
5.3.	Evolution of developmental robustness	83
6.	Conclusions	89
	References	91
3.	THE CIRCLE THAT NEVER ENDS: CAN COMPLEXITY BE MADE SIMPLE?	97
	<i>Donald C. Mikulecky</i>	
1.	Introduction: The Nature of the Problem and Why it Has No Clear Solution	97
1.1.	The human mind and the external world	99
1.2.	Science and the myth of objectivity	100
1.3.	Context dependence and self reference	102
2.	An Introduction to Relational Systems Theory	103
2.1.	Relational block diagrams	103
2.2.	Information as an interrogative. The answer to “why?”	104
2.3.	Functional components and their central role in complex systems	106
2.4.	The answer to “why is the whole more than the sum of its parts?”	106
2.5.	Reductionism and relational systems theory compared	107

2.6.	The functional component is not computable	108
2.7.	An example: the [M,R] system and the organism/machine distinction	108
2.8.	Relational models of mechanisms	112
2.9.	Newtonian dynamics is not unique; there are alternatives that yield equivalent results	112
2.10.	Topology, thermodynamics and relational modeling	114
2.11.	The mathematics of science or is all mathematics scientific?	117
2.12.	The parallels between vector calculus and topology . .	118
3.	The Structure of Network Thermodynamics as Formalism . .	118
3.1.	Network thermodynamic modeling is analogous to modeling electric circuits	119
3.2.	The network thermodynamic model of a system . . .	120
3.3.	Characterizing the networks using an abstraction of the network elements	120
3.4.	The nature of the analog models that constitute network thermodynamics	121
3.5.	The constitutive laws for all physical systems are analogous to the constitutive laws for electrical networks or can be constructed as the models for electronic elements	122
3.6.	The resistance as a general systems element	123
3.7.	The capacitance as a general systems element	124
3.8.	The topology of a network	126
3.9.	The formal description of a network	126
3.10.	The formal solution of a linear resistive network . . .	128
3.11.	The use of multiports for coupled processes: the entry to biological applications	130
3.12.	Linear multiports are based on non-equilibrium thermodynamics	130
4.	Simulation of Non-Linear Networks on Spice	133
4.1.	Simulation of chemical reaction networks	134
4.2.	Simulation of mass transport in compartmental systems and bulk flow	134
4.3.	Network thermodynamics contributions to theory: some fundamentals	135
4.4.	The canonical representation of linear non-equilibrium systems, the metric structure of thermodynamics, and the energetic analysis of coupled systems	135

4.5.	Tellegen's theorem and the onsager reciprocal relations (ORR)	136
5.	Relational Networks and Beyond	138
5.1.	A message from network theory	138
5.2.	An "emergent" property of the 2-port current divider	139
5.3.	The use of relational systems theory in chemistry and biology: past, present, and future	141
5.4.	Conclusion: there is no conclusion	144
	References	148
4.	GRAPHS AS MODELS OF LARGE-SCALE BIOCHEMICAL ORGANIZATION	155
	<i>Pau Fernández and Ricard V. Solé</i>	
1.	Introduction	155
2.	Basic Properties of Random Graphs	157
2.1.	Degree distribution	158
2.2.	Components	159
2.3.	Average path length	159
2.4.	Clustering	161
2.5.	Small-worlds	162
3.	Protein Structure and Contact Graphs	164
3.1.	Proteins are small worlds	165
3.2.	Hierarchical clustering in contact maps	166
4.	Protein Interaction Networks	169
4.1.	Assortativeness and correlations	171
4.2.	Correlation profiles	172
4.3.	Proteome model	175
5.	Gene Networks	180
6.	Overview	187
	Acknowledgements	188
	References	188
5.	QUANTITATIVE MEASURES OF NETWORK COMPLEXITY	191
	<i>Danail Bonchev and Gregory A. Buck</i>	
1.	Some History	191
2.	Networks as Graphs	193
2.1.	Basic notions in graph theory [36-38]	193
2.2.	Adjacency matrix and related graph descriptors	195
2.3.	Cluster coefficient and extended connectivity	196

2.4.	Graph distances	198
2.5.	Weighted graphs	201
3.	How to Measure Network Complexity	202
3.1.	Careful with symmetry!	202
3.2.	Can Shannon's information content measure topological complexity?	203
3.3.	Global, average, and normalized complexity	205
3.4.	The subgraph count, SC , and its components	207
3.5.	Overall connectivity, OC	210
3.6.	The total walk count, TWC	211
4.	Combined Complexity Measures Based on the Graph Adjacency and Distance	213
4.1.	The A/D index	213
4.2.	The complexity index B	215
5.	Vertex Accessibility and Complexity of Directed Graphs	218
6.	Complexity Estimates of Biological and Ecological Networks	221
6.1.	Networks of Protein Complexes	222
6.2.	Food webs	226
7.	Overview	230
	Acknowledgement	232
	References	232
6.	CELLULAR AUTOMATA MODELS OF COMPLEX BIOCHEMICAL SYSTEMS	237
	<i>Lemont B. Kier and Tarynn M. Witten</i>	
1.	Reality, Systems, and Models	237
1.1.	Introduction	237
1.2.	The "what" of modeling and simulation	238
1.3.	Back to models	244
1.4.	Models in chemistry and molecular biology	246
2.	General Principles of Complexity	248
2.1.	Defining complexity: complicated vs. complex	248
2.2.	Defining complexity: agents, hierarchy, self-organization, emergence, and dissolution	250
3.	Modeling Emergence in Complex Biosystems	257
3.1.	Cellular automata	257
3.2.	The general structure	258
3.3.	Cell movement	262
3.4.	Movement (transition) rules	267
3.5.	Collection of data	273

4.	Examples of Cellular Automata Models	274
4.1.	Introduction	274
4.2.	Water structure	275
4.3.	Cellular automata models of molecular bond interactions	277
4.4.	Diffusion in water	280
4.5.	Chreode theory of diffusion in water	283
4.6.	Modeling biochemical networks	289
5.	General Summary	297
	References	298
7.	THE COMPLEX NATURE OF ECODYNAMICS	303
	<i>Robert E. Ulanowicz</i>	
1.	Introduction	303
2.	Measuring The Effects of Incorporated Constraints	306
3.	Ecosystems and Contingency	307
4.	Autocatalysis and Non-Mechanical Behavior	311
5.	Causality Reconsidered	316
6.	Quantifying Constraint in Ecosystems	318
7.	New Constraints to Help Focus a New Perspective	324
	Acknowledgements	327
	References	327
	NAME INDEX	331
	SUBJECT INDEX	337

Chapter 1

ON THE COMPLEXITY OF FULLERENES AND NANOTUBES

Milan Randić

National Institute of Chemistry, Ljubljana, Slovenia

Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311

Xiaofeng Guo

Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311

Department of Mathematics, Xiamen University, Xiamen Fujian 361005, P. R. China

Dejan Plavšić

Institute Rudjer Bošković, Zagreb, 41001 Croatia

Alexandru T. Balaban

Texas A & M University at Galveston, Galveston, TX 77551

1. Introduction

“As it is usually rather difficult to produce good definitions for very general concepts, we have to let examples guide us on our way.”

Hans Primas [1]

The topic of molecular complexity continues to attract attention of chemists as is reflected by recent literature on this topic [2-7]. First we should mention that there are two distinct aspects of the notion of complexity measures: (i) the complexity relating to chemical reaction and synthesis of compounds, pioneered by Bertz [8-10], and the complexity of an isolated structure or molecular graph. We will relate in this contribution to molecular graphs. Even a superficial browsing through the literature immediately shows that different authors used different structural invariants as indices of molecular complexity. Among the invariants used we find: the Wiener index [11], the count of spanning trees [12-14], the count of paths [15], the count of walks [7], and quantities based on augmented valence [2,3]. For an informative introduction to various problems relating to the notion of complexity of graph we refer readers to a paper by Bonchev [5], where brief

historic developments were outlined. The first paper on graph complexity has been published about 35 years ago by Mowshovitz [16], who related complexity of graphs to entropy. On the other hand, Shannon's formula for an "information content" of objects offers an alternative approach to the concept of complexity [17]. Although the two approaches appear different, they have as common underlying features randomness and order. Patterns of high order and high symmetry, showing structural "uniformity" will be associated with low entropy and low information content, while patterns exhibiting little symmetry and no apparent regularities, appearing as random, will show high entropy and high information content, because local features will have distinct characteristics.

Most authors appear not to have considered a rigorous definition of complexity but instead considered relative measures of complexity. In this respect complexity remains as one of those apparently useful, but with respect to the rigor of its definition, elusive chemical concepts to which Hans Primas alludes in the opening quote to this article. That the notion of complexity is inherently complex is reflected already by a list of desirable attributes for a complexity index as specified by Bonchev and Polansky [18] and reproduced in Table 1.1. Most authors appear to agree that the complexity of mathematical objects such as chemical structures increases with information content, molecular size, connectivity of the graph, molecular branching, cyclicity of molecules, multiplicity of bonds, and the presence of heteroatoms (coloring of graphs), while it should decrease with increasing

Table 1.1. Desirable properties for complexity index (after Bonchev and Polansky [18])

No.	Complexity measures should:
1	Be independent of the nature of the system
2	Be specified within a unique theoretical conception
3	Take into account the different complexity levels and their hierarchy
4	Exhibit stronger dependence on relations between the elements than on element numbers
5	Increase monotonically with the number of different complexity features
6	Agree with the intuitive idea of complexity
7	Differentiate the nonisomorphic systems
8	Not to be too sophisticated
9	Be applicable to practical purposes

symmetry properties of objects under consideration. Again one may observe that most of the above listed qualities of mathematical objects themselves remain somewhat ambiguous, either lacking rigorous definition, or having alternative and arbitrary interpretations as to how to select the descriptor to be used for characterization of structures. The degree of complexity will depend on which structural property will be selected for characterizing the given objects. In the case of molecular graphs, for instance, one can consider equivalence classes (that is symmetry), path distributions, degree distributions, edge types, etc., of molecular graphs, and each time a different relative measure of complexity may follow.

2. On the Complexity of the Complexity Concept

“Every theoretical framework must start with some undefined concepts which need no further explanation and can be taken as granted without proof, definition or analysis. Such concepts are called primitive concepts.”

Hans Primas [19]

Before we proceed, we need to justify the statement that “the complexity is inherently complex,” because we are “explaining” a concept by itself! Observe the same with the list of desirable attributes of the notion of complexity of the Table 1.1, where three out of the nine “desiderata” referring to complexity are using the term complexity! So it appears that we are trapped in “circular reasoning,” not an uncommon use of faulty logic. Roald Hoffmann has written a paper “*Nearly Circular Reasoning*” [20] in which he argued that most chemists have no training in logic and therefore may consider some illogical alternatives. Then he continued to point out that, though ignorant of logic, no chemist has tried to repudiate logic, so that statements that may formally appear illogical could nevertheless be useful. In this respect the notion of complexity is not an isolated case of “circular reasoning.” Consider for instance the notion of aromaticity [21], which often has been “clarified” by using as criteria of aromaticity selective properties of “aromatic compounds.” However, that presumes that we already know which compounds are aromatic to start with, without stating how one defines an aromatic compound. Thus again we define aromaticity in terms of aromatic compounds! The problem of aromaticity, at least for polycyclic conjugated hydrocarbons, has been solved [21] (at least for those

willing to accept the offered solution) so that any particular confusing use of logic can be avoided. The concept of complexity, however, remains to haunt us and to present challenges that continue to stay open. The source of the difficulty may well be the lack of “primitive” concepts relating to complexity, which once they are identified could possibly clarify much of the current “fuzziness” of the notion of complexity. The case of “aromaticity” is in this respect a valid illustration, which became “tamed” once the notion of “conjugated circuits” [22, 23] was taken as a “primitive” of aromaticity.

3. Complexity and Branching

Of the nine desirable attributes for the concept of complexity listed in Table 1.1 perhaps the closest to an overall characterization of complexity is the requirement that any complexity measure should “exhibit stronger dependence on the relations between the elements than on the element number.” Thus the *pattern* in which components of a system are related appears more important than the *number of components*. We use the term “components” in a broad sense as “*components of a system*” and not in a narrow sense as “*components of a graph*.” Components of a graph are defined as any subset of vertices and incident edges, taking only edges between vertices in the subset [24]. One is tempted than to consider *components* and *pattern* as potential “primitives” of complexity, except that both concepts remain vague until specified in each application.

In view of these difficulties it appears that in the case of graphs one should start with *connectivity* and consider well-defined quantities based on the concept itself. Although connectivity does not qualify as elementary “primitive” by being a global rather than local graph property, connectivity can be rigorously defined. Mathematically connectivity is reflected in the binary graph adjacency matrix **A** [24]. From the adjacency matrix one can construct a number of graph invariants such as the graph eigenvalues and the characteristic polynomial. In addition, we may consider various topological indices [25-27], (in particular the connectivity index [28, 29], which has been interpreted as a measure of molecular branching). We should also mention the extended connectivity [30-33] and walks of different length, which have been considered by Rucker and Rucker as useful measures of graph complexity [34]. Walks in a graph have, beside the simple structural meaning not requiring any explanation whatsoever, a simple relationship

to the adjacency matrix $\mathbf{A}$ in that they can be “counted” by simply raising $\mathbf{A}$ to higher powers. Moreover, although apparently their number increases without limit with their length, in fact because of the Cayley-Hamilton theorem [35], which states that matrix $\mathbf{A}$ satisfies its own characteristic polynomial, the number of walks that carry distinct informational content is finite. On the other hand, the extended connectivity, which is calculated iteratively by augmenting the valence of vertices by the number of k nearest neighbors, then next to nearest, etc., in the limit of $k \rightarrow \infty$ where k is the exponent of the matrix $\mathbf{A}^k$, yields to the leading eigenvalue of the adjacency matrix [36]. Although the mentioned invariants can be considered as alternative indices of complexity, some of them have found alternative interpretations. Thus in the case of trees (acyclic graphs) Lovasz and Pelikan have interpreted the leading eigenvalue of the adjacency matrix as an index of molecular branching [37]. Hence, are we in this way using branching indices for measuring complexity and in the case of acyclic graphs equating branching with complexity?

Apparently the answer may appear trivial, if one accepts that “another name of branching” is “acyclic complexity” [38]. In other words, one may ask the following question. “Is for acyclic systems ‘complexity’ identical to ‘branching’?” However, if one is to answer the question, one should have a valid and generally accepted definition of ‘complexity’ and similarly valid and generally accepted definitions of ‘branching’. We have neither! Recall that some authors (e. g. Ruch and Gutman [39]) have in a way challenged the notion of quantifying branching by arguing that ‘branching’ can only be defined by partial ordering. This then implies that there are structures which neither dominate each other nor are dominated by each other with respect to branching. An implication of this attitude is that, at least in the case of acyclic graphs, the same holds for complexity! In other words, while for some structures we can establish “complexity dominance”, there are structures (acyclic graphs) for which we could not determine their relative complexity.

Perhaps we are half-way to resolving all these difficulties because relatively recently an ingenious approach [40-41] (if one can refer thus to one’s own work) has been suggested for a quantitative characterization of molecular branching that is devoid of “arbitrary” decisions and choices (which characterize more or less all “branching indices”, starting with the suggestion of Lovasz and Pelikan [37] that the leading eigenvalue of the adjacency matrix of acyclic graphs can be a measure of their branching).

According to this novel view on branching, which at least in the case of smaller molecules appears satisfactory, all eigenvalues of an acyclic graph are expressed as linear combinations of eigenvalues of linear path graphs P_n and star graphs $K_{1,n}$. The coefficients of such a linear combination give the “degree” of mixing of the linear graph and the star graph, which then is interpreted as the “degree of branching.” It remains to be seen whether in the case of larger graphs a similar analysis will suffice or one will have to introduce additional graphs to augment the assumed “basis” of P_n and $K_{1,n}$.

As we mentioned, to make things “worse”, one has to recognize that there have been also some ambiguities concerning the definition of molecular branching. Besides having a number of alternative “branching indices”, a question has been raised whether one can differentiate between structures that have the same number of branches, such as for example 2-methylheptane, 3-methylheptane and 4-methylheptane. According to Ruch and Gutman [39], the concept of branching should not be extended beyond partitioning of vertices of graph according to their degrees. A similar “cautious” attitude appears to have been suggested for characterizing the elusive aromaticity of benzenoid hydrocarbons [42, 43], but most chemists have been searching for an “ideal” characterization of aromaticity anyway. Although we may have to wait for a while for consensus, it is not easy to see how will be refuted the recently presented arguments in favor of the characterization of aromaticity in terms of the $(4n + 2)$ and the $(4n)$ conjugated circuits contained in individual Kekulé valence structures of such molecules [44]! Similarly we have to wait to see how well will the novel characterization of branching be accepted or amended, but be it as it may, we appear closer to having a plausible definition of branching – while awaiting a similar clarification of complexity even in the case of acyclic structures.

Let us return to the issue of branching. Without a precise definition of “molecular branching” it is not easy to argue for the necessity to differentiate among structures having the same vertex degree distribution and characterize some of them as more branched than others. Nevertheless, the recently proposed scheme based on representing the leading eigenvalue of a graph as a linear combination of the leading eigenvalues of the ‘extreme’ graphs (the linear and the star graph) points convincingly to 2-methylheptane to be less “branched” than 3-methylheptane and 4-methylheptane. It is self-evident that among isomers a “linear structure”

(n-alkane) is the least branched while the “star graph” in which all vertices are neighbors to a single central vertex is the most branched. By expanding the eigenvalues of a “graph in-between” one obtains immediately a quantitative estimate of the degree of branching of such a graph [40]. When this is applied to 2-methylhexptane (2-M), 3-methylheptane (3-M), and 4-methylheptane (4-M) one finds that for 2-methylheptane the coefficients of the expansion point to 2-methylheptane to be closer to n-octane (the linear case) than are 3-methylheptane and 4-methylheptane, which have larger coefficients of the “star graph” $K_{1,7}$. We may add that the relative ordering of 2-M, 3-M, 4-M agrees with the ordering of these three isomers of octane by numerous topological indices (including the Wiener index), which have been considered as indices that “parallel” complexity. But there are indices that reverse the relative order of 2-M and 3-M and if they would be used as indices of complexity one would have to conclude that 2-M-alkanes are more complex than 3-M-alkanes! Two of these indices which reverse the order of 2-M and 3-M are the connectivity index χ and the Hosoya index Z , both showing good correlations with numerous molecular properties. And one should recall that even though the majority of molecular graph theoretical indices lead to the order: n-alkane > 2-M-alkane > 3-M-alkane, some physical-chemical properties (such as the boiling points) lead to the relative ordering: n-alkane > 3-M alkane > 2-M alkane, which is one of the reasons why the connectivity index and Hosoya’s Z index outperform many other topological indices in simple regressions of numerous physical-chemical properties [45].

4. Complexity of Smaller Molecules

Before discussing the complexity of fullerenes let us briefly discuss the complexity of smaller molecules. Several comparisons between topological indices have been recently published [46]. Not long ago, Nikolić, Trinajstić, Tolić, Rücker and Rücker [46a] compared complexities of several sets of smaller structures as calculated by a collection of 17 topological indices, which included complexities based on the leading eigenvalue of the adjacency matrix (λ_1) and complexities as given by the number of walks. We reproduce in Table 1.2 their results for the five isomers of hexane but only for the complexities based on the leading eigenvalue of the adjacency matrix and complexity as given by the number of walks:

Table 1.2. Complexities of five isomers of hexane as calculated for the complexities based on the leading eigenvalue of the adjacency matrix and complexity as given by the number of walks

Hexane isomer	walks	λ_1
n-hexane	222	1.8019
2-methylpentane	268	1.9021
3-methylpentane	284	1.9319
2,3-dimethylbutane	330	2.0000
2,2-dimethylbutane	370	2.0743

One can observe the complete parallelism between the two alternative characterizations of complexity, namely the count of walks and the leading eigenvalue. There is no doubt that a similar parallelism will be found for other pairs of topological descriptors just as there will be descriptors that will show different behavior. This may tend to strengthen the impression that the selected descriptors are good indices of the elusive complexity. However, we want to add a neglected aspect showing that there is a need to look at larger systems and see what happens there.

In Table 1.3 we have listed for the 18 isomers of octane (shown in Fig. 1.1), the count of walks in these molecules, the leading eigenvalues of the adjacency matrix, the reciprocal of the leading values of the path matrix (normalized by multiplying the reciprocal by n , the number of carbon atoms), which may be viewed as an alternative to the “branching index”), the values of the complexity based on the augmented valence [2, 3], and a closely related index based on “augmented” path counts [47]. The path matrix is defined first by constructing a graphical matrix [48, 49] in which matrix elements are defined by the subgraph representing the number of paths between vertices i and j , which is subsequently transformed into a numerical matrix by replacing the path subgraphs with the maximum eigenvalue of those paths. In Table 1.4 we have illustrated the path matrix on 3-methylheptane (assuming conventional numbering of carbon atoms). In the top part we show the matrix elements as paths of different length p_i , while in the lower part of Table 1.4 we show the numerical matrix in which each p_i element is replaced by the corresponding leading eigenvalue of the path of length i .

The leading eigenvalue of the path matrix offers an alternative branching index, which has an important advantage over the leading eigenvalue of the

Table 1.3. The count of walks, the leading eigenvalue λ_1 of the adjacency matrix **A**, the index derived from the leading eigenvalue of the path matrix **P** and the augmented valence complexity index ($\xi\xi$) for the 18 constitutional isomers of octane

Octane isomer	walks	λ_1 of A	n/λ_1 of P	$\xi\xi$
n-octane	627	1.8794	0.76937	36.047
2-methylheptane	764	1.9499	0.77771	37.500
3-methylheptane	838	1.9890	0.78156	38.156
4-methylheptane	856	2.0000	0.78269	38.344
2,5-dimethylhexane	911	2.0000	0.78653	39.000
3-ethylhexane	928	2.0285	0.78666	39.000
2,4-dimethylhexane	997	2.0421	0.79090	39.750
2,3-dimethylhexane	1068	2.0743	0.79286	40.125
3,4-dimethylhexane	1136	2.0953	0.79615	40.688
2,2-dimethylhexane	1142	2.1120	0.79352	40.313
2-methyl-3-ethylpentane	1152	2.1010	0.79738	40.875
2,3,4-trimethylpentane	1296	2.1358	0.80381	42.000
3,3-dimethylhexane	1301	2.1566	0.80005	41.438
2,2,4-trimethylpentane	1317	2.1490	0.80367	42.000
3-methyl-3-ethylpentane	1441	2.1889	0.80553	43.375
2,2,3-trimethylpentane	1536	2.2060	0.80995	43.125
2,3,3-trimethylpentane	1609	2.2216	0.81208	43.000
2,2,3,3-tetramethylbutane	2047	2.3028	0.82527	45.750

adjacency matrix (the index proposed by Lovasz and Pelikan). Apparently the leading eigenvalue of the path matrix shows low degeneracy, if any, because so far no case of graphs having the same leading eigenvalue of the path matrix has been detected. However, one should not be surprised if a systematic search eventually will detect pairs of degenerate graphs with respect to the path matrix, though low degeneracy may be expected on the grounds that while the characteristic polynomial of the adjacency matrix has integers as coefficients, this is not the case with the path matrix, the coefficients of the characteristic polynomial of which could be real irrational numbers.

Let us point out that all the four “indices” displayed in Table 1.3 show considerable parallelism. Because the count of walks has been generally accepted as a *bona fide* descriptor of complexity and the same holds for the leading eigenvalue of the adjacency matrix, we may claim that the leading eigenvalue of the path matrix and the column shown as $\xi\xi$ index based on augmented valence can equally measure the level of complexity. This claim

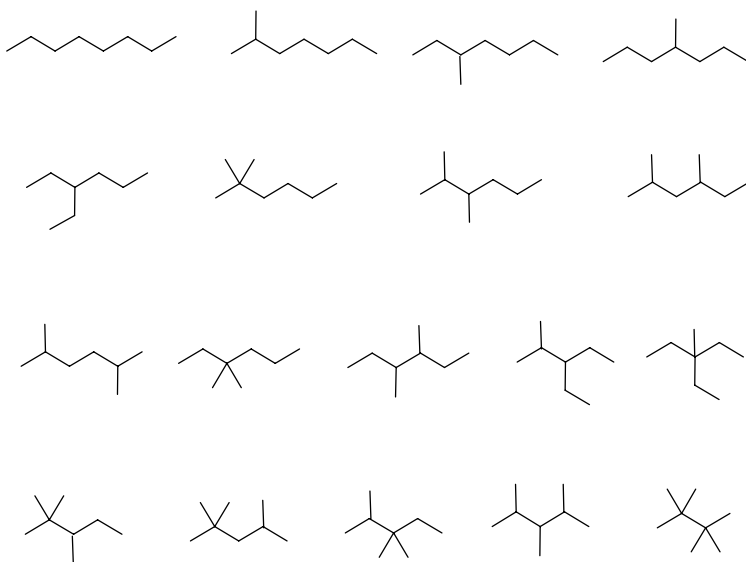


Figure 1.1. The 18 isomers of octane.

is further substantiated by Figs. 1.2, 1.3 and 1.4 in which we show partial orders for the 18 isomers of octane based on the count of walks and the leading eigenvalue of the adjacency matrix (Fig. 1.2), the count of walks and the leading eigenvalue of the path matrix (Fig. 1.3), and finally the count of walks and the $\xi\xi$ index (Fig. 1.4). Observe that the Hasse diagrams which depict the three partial orders have a similar form, ordering identically the lowest eight isomers, but showing occasional branching for octane isomers for which the two measures of complexity disagree.

Table 1.4. The path matrix $\mathbf{P}$ for 3-methylheptane in symbolic form followed by numerical data

	1	2	3	4	5	6	7	8
1	0	p_1	p_2	p_3	p_4	p_5	p_6	p_3
2		0	p_1	p_2	p_3	p_4	p_5	p_2
3			0	p_1	p_2	p_3	p_4	p_1
4				0	p_1	p_2	p_3	p_2
5					0	p_1	p_2	p_3
6						0	p_1	p_4
7							0	p_5
8								0

	1	2	3	4	5	6	7	8
1	0	1.0000	1.4142	1.6180	1.7321	1.8019	1.8478	1.6180
2		0	1.0000	1.4142	1.6180	1.7321	1.8019	1.4142
3			0	1.0000	1.4142	1.6180	1.7321	1.0000
4				0	1.0000	1.4142	1.6180	1.4142
5					0	1.0000	1.4142	1.6180
6						0	1.0000	1.7321
7							0	1.8020
8								0

Since there are other complexity indices, one may expect questions relating to novelty of additional complexity indices. In this respect we feel that at least two aspects when comparing different indices have to be considered: (1) Do new indices suggest different ordering among structures of the same size (measured by the number of atoms in a molecule in the case of molecular graphs)? (2) Do new indices offer alternative structural interpretations for the proposed complexity measure, and how do they differ from other structural interpretations of molecular descriptors? Consider for example viewing the leading eigenvalue of $\mathbf{A}$ as a complexity index. This leading eigenvalue has already been suggested as a measure of the degree of branching of molecular skeletons, which makes their interpretation as a measure of complexity somewhat overlapping with that of branching, possibly causing confusion rather than clarification. In the case of Table 1.3 one may insist that all listed indices characterize the degree of complexity, because the extreme (top and bottom) isomers apparently correspond to intuitive ideas of complexity, at least as outlined by Polansky and Bonchev [18]. However, at the same time and with the same certainty, one may argue that the indices represent the elusive “degree of branching.” Again the isomers at both extremes of the scale, *n*-octane and 2,2,3,3-tetramethylbutane, fully agree with our intuitive perception of branching. So is there any difference between complexity and branching when dealing with acyclic structures?

One should recall that Bonchev (after having pioneered the introduction of information theory into new molecular descriptors) has also pioneered various quantitative measurements of complexity in chemistry, as attested by the rich bibliography [50].

The following question can be raised (in fact the question has been raised by a referee of a related article). Do the three indices n/λ_1 , $\xi\xi$ and π add any novelty, and in particular, do they add any new conclusions to those of

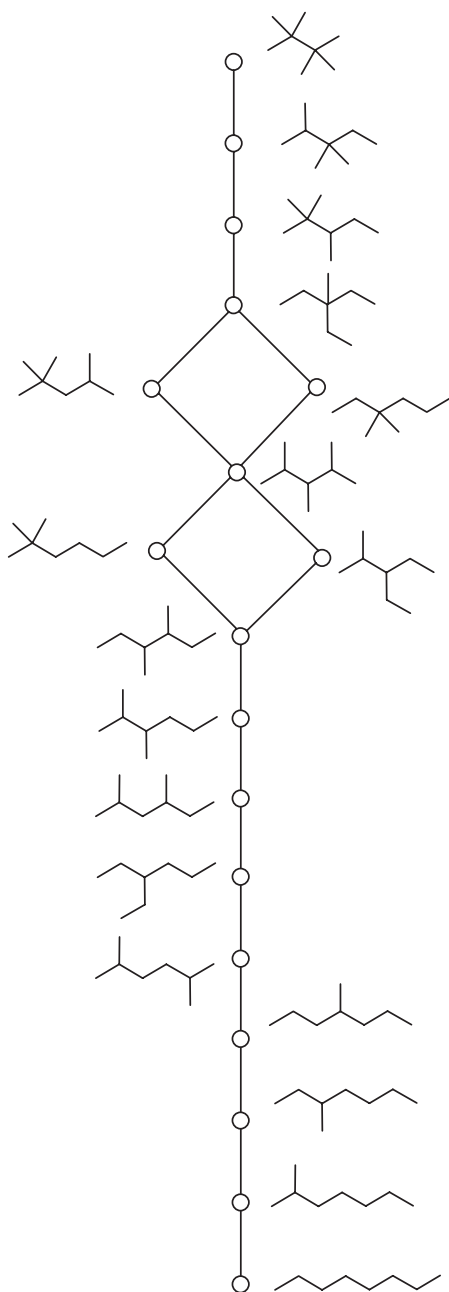


Figure 1.2. Partial order for 18 isomers of octane based on the count of random walks and the magnitude of the leading eigenvalue of the adjacency matrix.

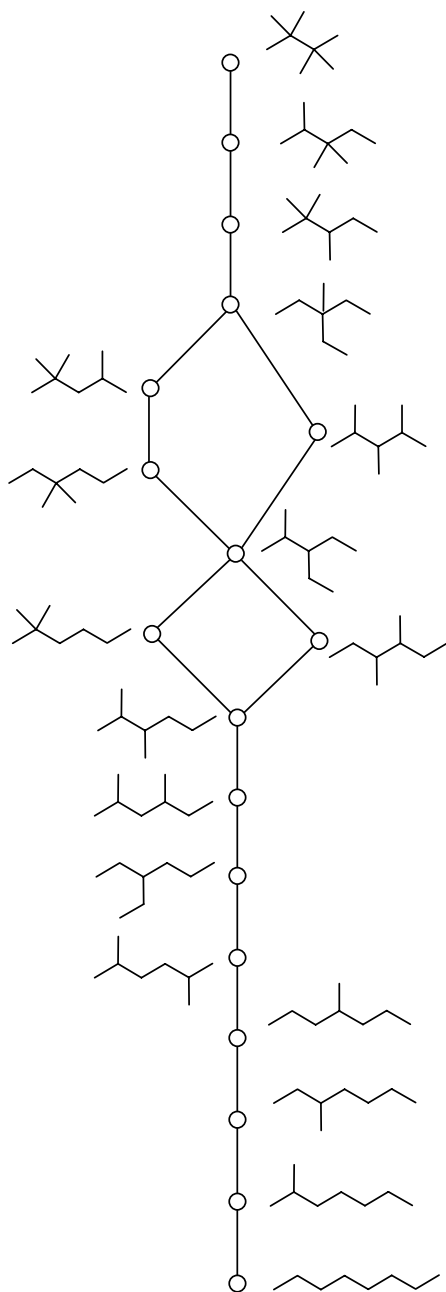


Figure 1.3. Partial order for 18 isomers of octane based on the count of random walks and the magnitude of the leading eigenvalue of the path matrix.

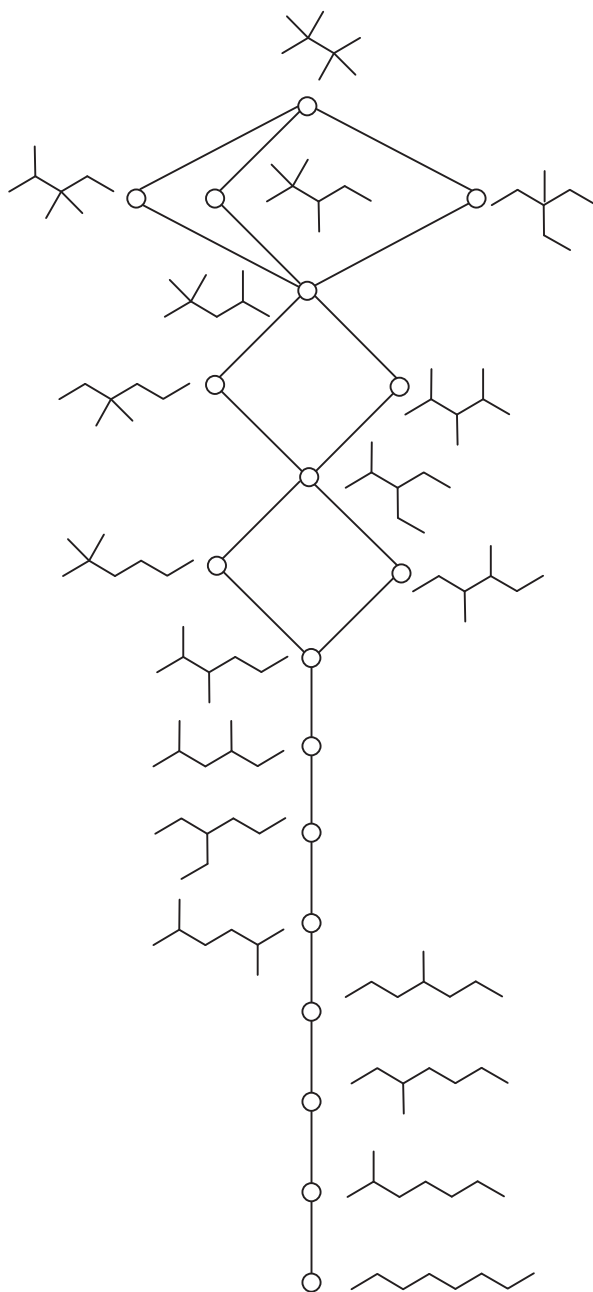


Figure 1.4. Partial order for 18 isomers of octane based on the count of random walks and the magnitude of the augmented valence complexity index.

Nikolić *et al.* [46a] who examined 17 topological indices on 33 smaller molecules graphs that include linear alkanes for $n = 2$ to $n = 10$, all hexane isomers, cycles depicting carbon skeletons of monocycloalkanes with $n = 3$ to $n = 10$, and five polycyclic $(CH)_8$ isomers of cubane? The answer is yes!

First, let us point to major conclusions of the review article on complexity by Nikolić *et al.* [46a] who at the end of their article summarized their observations. One of their conclusions refers to random walks (the total walk count in their terminology) and the leading eigenvalue of the adjacency matrix. According to these authors: *“The total walk count and the leading eigenvalue of the adjacency matrix cannot discriminate between isomeric regular graphs.”* This means that if the walk count or the leading eigenvalue is taken as a measure of complexity, all isomeric regular graphs are equally complex. That limitation need not necessarily eliminate the count of walks and the leading eigenvalues as complexity indicators, but rather shows their limited domain of application. The desideratum for the complexity measures of Bonchev and Polansky listed in Table 1.1 (as the condition # 7) suggests that non-isomorphic graphs should have different complexity values; while this is a desirable recommendation, it does not have to be a priori viewed as a necessary condition, particularly when one tries to characterize a quantity that is at best vaguely known.

More interesting for the characterization of complexity is to see if various descriptors agree and parallel each other or show different “scales” of complexity, that is, different ordering for the same set of compounds. In this respect the total count of walks and the leading eigenvalue of the adjacency matrix as reported on hexane isomers by Nikolić *et al.* [46a] fully agree. Their result may thus give the impression that this will hold even for larger systems, but as we can see from our Table 1.3 and Fig. 1.2, this is not the case. Two pairs from the 18 isomers of octane do not follow this parallelism, namely the pair: 2-methyl-3-ethylpentane and 2,2-dimethylhexane, and the pair: 2,2,4-trimethylpentane and 3,3-dimethylhexane. We think that it is important to point to such disagreements and not to argue which of the two alternative descriptors is “better” as a descriptor of molecular complexity, because both may be equally good or equally bad. Instead, when considering complexity, (1) one should focus attention on larger and more complex structures and (2) as much as possible, one should eliminate from the “competition for the characterization of complexity” those descriptors which show “excessive” degeneracy. In Table 1.5 we have listed brief

Table 1.5. Classification of molecular descriptors with respect to their power to discriminate non-isomorphic graphs

Highly degenerate	Moderate	Limited degeneracy
Number of kinds of connected subgraphs Zagreb indices Wiener's W index Hosoya's Z Index The connectivity index	The leading eigenvalue of adjacency matrix Balaban's J index Path/Walk indices	The leading eigenvalue of path matrix Molecular ID number Prime number ID Extended Adjacency matrix Topological Index (EATI) Balaban's BID Xu index

descriptors with excessive, moderate and limited degeneracy – a rather brief list which is clearly subjective, at least when it comes to classification of descriptors that can represent the “borderline” cases. Nevertheless, by eliminating most of highly degenerate and possibly even some of the moderately degenerate descriptors, we may be closer to identifying structural features that are important for complexity.

Before turning to the problem of complexity of fullerenes let us express our opinion that considerations relating to molecular branching and its numerical characterization ought to be extended to other often mentioned “complexity” attributes, such as the connectivity, the cyclicity, the unsaturation, the presence and role of heteroatoms, etc. In this way we hope that we will arrive at better insights as to what really the term complexity represents, or ought to represent, and how it differs from other structural concepts used for characterizing the complexity (that is, connectivity, branching, cyclicity, etc.).

5. Augmented Valence as a Complexity Index

We have already mentioned that on one hand the count of walks in molecular graphs, and on the other hand the leading eigenvalue of the adjacency matrix **A**, both offer insights into molecular complexity. That these two apparently conceptually different quantities qualify for a characterization of the same structural concept – that of complexity – need not be so

surprising in view of the fact that there are connections between the two, even if they are not obvious. As it is well known, the elements of various powers of the adjacency matrix count random walks [24].

The concept of “extended connectivity”, which forms the basis of one of the earliest algorithms for unique labeling of atoms in molecules, due to Morgan [30], is also related to the higher powers of the adjacency matrix A in that the i -th iteration of Morgan algorithm can be deduced directly from the i -th power of A [31]. The extended connectivity is a property of individual atoms and is given by the sequence of entries obtained by adding to the initial value of the valence of an atom the valences of adjacent atoms, followed by those of atoms once removed, . . . to k -removed ones. We illustrate in Fig. 1.5 the calculation of the extended connectivity for carbon atoms of 3-methylheptane, which among all octane isomers is the only “identity tree” because its only symmetry is the identity operation. In the lower part of Table 1.6 we have listed the extended connectivity for carbon atoms of 3-methylheptane. As one can see, the extended connectivity values increase with each iterative step, which takes into consideration indirectly the valences of more distant carbon atoms. The augmented valence, based on summing the valences of neighbors at larger and larger distances from a given atom, shown in the upper part of Table 5, also incorporates information on more distant neighbors. However with factor $1/2^k$ the role of more distant neighbors is gradually attenuated. In order to reduce the unlimited growth of extended connectivity with each successive iteration we may introduce as a normalization factor the reciprocal factorial $1/k!$, which will curb the growth of extended connectivity even faster than $1/2^k$. If we now combine the contributions from this “controlled” extended connectivity as illustrated in the lower part of Table 5 when contributions are limited to the first $k = 6$ steps, we obtain a similar local characterization of atom complexity as we did with the augmented valence. Observe that all the relative magnitudes are the same whether we use the augmented valence or the suitably normalized extended connectivity – which points to an inherent robustness of the proposed measures of local atom complexity.

The idea of using the inverse power $1/2^k$ weights originated with the wish to diminish the role of more distant neighbors on contributions to local molecular complexity. The selection of the weighting factors $1/2^k$ is to a degree arbitrary and other functions that decrease the role of more distant neighbors are possible. To what extent various alternative weighting functions may influence the relative values of complexity indices has yet to

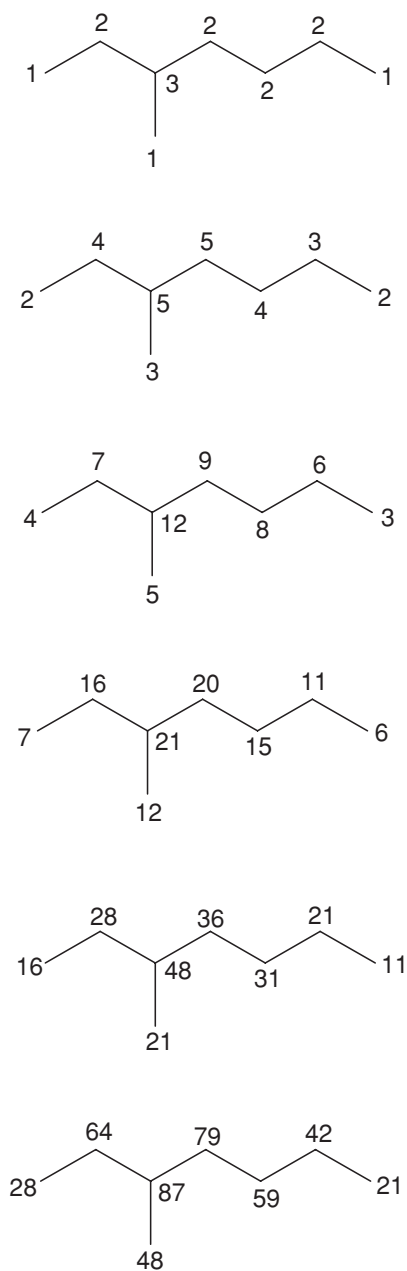


Figure 1.5. Illustration of several initial steps in the construction of extended connectivity for 3-methylheptane.

Table 1.6. Atomic contributions for augmented valence complexity index for carbon atoms of 3-methylheptane (top part) and atomic contributions for augmented extended connectivity complexity index for carbon atoms of 3-methylheptane (bottom part)

Carbon atom	Neighbor valence sum	Augmented valence	Sum
1	1, 2, 3, 3, 2, 2, 1	$1 + 2/2 + 3/4 + 3/8 + 2/16 + 2/32 + 1/64$	3.328125
2	2, 4, 3, 2, 2, 1	$2 + 4/2 + 3/4 + 2/8 + 2/16 + 1/32$	5.156250
3	3, 5, 3, 2, 1	$3 + 5/2 + 3/4 + 2/8 + 1/16$	6.562500
4	2, 5, 5, 2	$2 + 5/2 + 5/4 + 2/8$	6.000000
5	2, 4, 4, 3, 1	$2 + 4/2 + 4/4 + 3/8 + 1/16$	5.437500
6	2, 3, 2, 3, 3, 1	$2 + 3/2 + 2/4 + 3/8 + 3/16 + 1/32$	4.593750
7	1, 2, 2, 2, 3, 3, 1	$1 + 2/2 + 2/4 + 2/8 + 3/16 + 3/32 + 1/64$	3.046875
8	1, 3, 4, 3, 2, 1,	$1 + 3/2 + 4/4 + 3/8 + 2/16 + 1/32$	4.031250

Carbon atom	Extended connectivity	Augmented extended connectivity	Sum
1	1, 2, 4, 7, 16, 28, . . .	$1 + 2/2! + 4/3! + 7/4! + 16/5! + 28/6!$	3.13056
2	2, 4, 7, 16, 28, 64, . . .	$2 + 4/2! + 7/3! + 16/4! + 28/5! + 64/6!$	6.15556
3	3, 5, 12, 21, 48, 87, . . .	$3 + 5/2! + 12/3! + 21/4! + 48/5! + 87/6!$	8.89583
4	2, 5, 9, 20, 36, 79, . . .	$2 + 5/2! + 9/3! + 20/4! + 36/5! + 79/6!$	7.24306
5	2, 4, 8, 15, 31, 57, . . .	$2 + 4/2! + 8/3! + 15/4! + 31/5! + 57/6!$	6.29583
6	2, 3, 6, 11, 21, 42, . . .	$2 + 3/2! + 6/3! + 11/4! + 21/5! + 42/6!$	5.19167
7	1, 2, 3, 6, 11, 21, . . .	$1 + 2/2! + 3/3! + 6/4! + 11/5! + 21/6!$	2.87083
8	1, 3, 5, 12, 21, 48, . . .	$1 + 3/2! + 5/3! + 12/4! + 21/5! + 48/6!$	4.07500

be investigated, but the choice selected here is, at least from a mathematical point of view, one of the simplest forms of a function that satisfies the assumed requirement for a local complexity measure to be less and less influenced by more and more distant neighbors.

For more details on the use of augmented valence as a measure of molecular complexity we direct readers to two introductory articles on this topic

published in *International Journal of Quantum Chemistry* [2] and *Croatica Chemica Acta* [3], and an article on application of augmented valence for characterization of complexity to transitive graphs representing degenerate rearrangements [51].

6. Complexity of Smaller Fullerenes

We will now examine twenty smaller fullerenes having from 20–60 carbon atoms, six of which are illustrated by Schlegel diagrams in Fig. 1.6 (C_{20} , C_{24} , C_{26} , C_{28} , C_{30} , and C_{32}). These are the same twenty fullerenes whose topological equivalence classes for carbon atoms and CC bonds were studied in some detail by Laidboeur, Cabrol-Bass and Ivanciuc [52], where illustrations of all twenty fullerenes can be found. In order to facilitate the comparison with the work of Laidboeur and collaborators, we have adopted here the same numbering for fullerenes and their carbon atoms as used in their paper [52]. In Table 1.7 we have listed the distance degree (DD) sequences (or DDSs) for all symmetry non-equivalent carbon atoms of all twenty fullerenes. The next column indicates the multiplicities (μ) of the

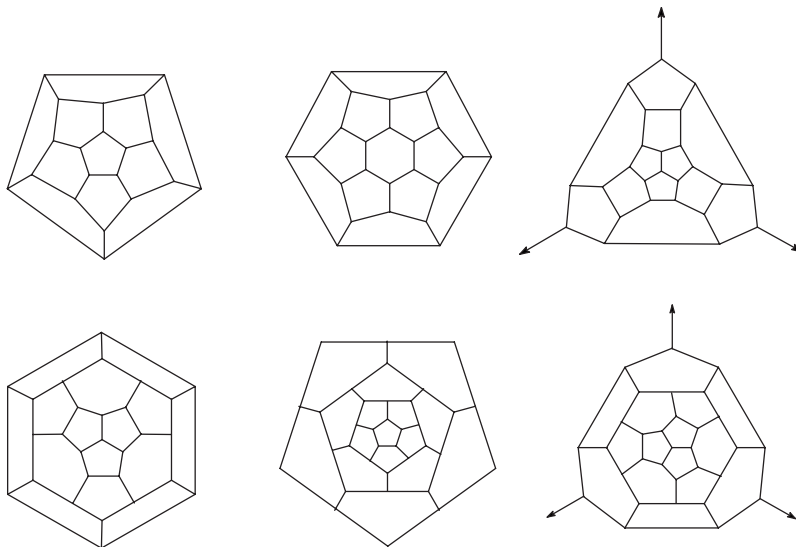


Figure 1.6. Several of the twenty smaller fullerenes having 20 through 60 carbon atoms.

Table 1.7. Distance degree sequences (DDSs), the multiplicity (μ), the atomic augmented complexity index ξ , the average atomic augmented complexity index ξ_{av} , and the overall atomic augmented complexity index $n \xi_{av}$ for twenty fullerenes C_n with $n = 20$ to 60 .

Fullerene	Atom	DD sequence	μ	ξ	ξ_{av}	$n \xi_{av}$
1. C_{20}	1	3, 9, 18, 18, 9, 3	20	14.9063	14.9063	298.1260
2. C_{24}	1	3, 9, 18, 21, 15, 6	12	15.7500	15.6094	374.6256
	7	3, 9, 18, 18, 15, 9	12	15.4688		
3. C_{26}	1	3, 9, 18, 18, 18, 9, 3	2	15.7031	15.9195	413.9079
	2	3, 9, 18, 18, 15, 15	6	15.6563		
	5	3, 9, 18, 21, 18, 9	12	16.0313		
	11	3, 9, 18, 21, 18, 9	6	16.0313		
4. C_{28}	1	3, 9, 18, 21, 21, 9, 3	12	16.2656	16.1987	453.5636
	2	3, 9, 18, 21, 18, 15	12	16.2188		
	7	3, 9, 18, 18, 18, 18	4	15.9375		
5. C_{30}	1	3, 9, 18, 18, 18, 18, 6	10	16.0313	16.4063	492.1890
	6	3, 9, 18, 21, 21, 15, 3	10	16.4531		
	11	3, 9, 18, 24, 21, 12, 3	10	16.7344		
6. C_{32}	1	3, 9, 18, 18, 18, 18, 9, 3	2	16.1016	16.6069	531.4208
	2	3, 9, 18, 18, 21, 18, 9	6	16.2656		
	5	3, 9, 18, 21, 21, 18, 6	6	16.5938		
	6	3, 9, 18, 21, 21, 18, 6	6	16.5938		
	11	3, 9, 18, 24, 24, 15, 3	6	17.0156		
	12	3, 9, 18, 21, 24, 18, 3	6	16.7344		
7. C_{34}	1	3, 9, 18, 27, 27, 18	1	17.6250	16.8144	571.6896
	2	3, 9, 18, 24, 27, 15, 6	3	17.2500		
	5	3, 9, 18, 21, 21, 21, 9	6	16.7344		
	7	3, 9, 18, 21, 21, 18, 12	3	16.6875		
	14	3, 9, 18, 18, 21, 21, 12	3	16.4063		
	17	3, 9, 18, 21, 24, 15, 9, 3	6	16.5939		
	18	3, 9, 18, 21, 21, 18, 12	6	16.6875		
	29	3, 9, 18, 24, 24, 18, 6	6	17.1563		
8. C_{36}	1	3, 9, 18, 21, 24, 21, 9, 3	12	16.9453	17.0391	613.4040
	5	3, 9, 18, 24, 24, 18, 9, 3	12	17.2266		
	7	3, 9, 18, 21, 24, 21, 9, 3	12	16.9453		
9. C_{38}	1	3, 9, 18, 18, 18, 18, 18, 9, 3	2	16.3008	17.1075	650.0850
	2	3, 9, 18, 18, 18, 18, 15, 15	6	16.2891		
	3	3, 9, 18, 21, 21, 18, 15, 9	12	16.8047		
	6	3, 9, 18, 24, 27, 24, 9	12	17.5781		
10. C_{38}	12	3, 9, 18, 27, 27, 21, 9	6	17.8594	17.2081	653.9078
	1	3, 9, 18, 18, 27, 27, 9, 3	1	16.9453		
	2	3, 9, 18, 21, 27, 24, 9, 3	3	17.2266		
	5	3, 9, 18, 24, 24, 21, 12, 3	6	17.3672		
	11	3, 9, 18, 24, 24, 18, 15, 3	6	17.3203		
	17	3, 9, 18, 21, 24, 21, 12, 6	3	17.0156		
	20	3, 9, 18, 21, 24, 21, 12, 6	6	17.0156		
						(cont.)

Table 1.7. Distance degree sequences (DDSs), the multiplicity (μ), the atomic augmented complexity index ξ , the average atomic augmented complexity index ξ_{av} , and the overall atomic augmented complexity index $n \xi_{av}$ for twenty fullerenes C_n with $n = 20$ to 60 . (Cont.)

Fullerene	Atom	DD sequence	μ	ξ	ξ_{av}	$n \xi_{av}$
11. C_{40}	26	3, 9, 18, 21, 24, 24, 12, 3	6	17.0859	17.3578	694.3128
	27	3, 9, 18, 21, 24, 21, 18	3	17.0625		
	32	3, 9, 18, 24, 27, 21, 9, 3	3	17.5078		
	34	3, 9, 18, 27, 27, 18, 9, 3	1	17.7891		
	1	3, 9, 18, 24, 24, 21, 15, 6	24	17.4375		
12. C_{40}	7	3, 9, 18, 18, 27, 27, 18	4	17.0625	17.3578	694.3120
	8	3, 9, 18, 21, 27, 24, 12, 6	12	17.2969		
	1	3, 9, 18, 21, 24, 21, 15, 9	6	17.0859		
	6	3, 9, 18, 21, 24, 24, 15, 6	6	17.1563		
	10	3, 9, 18, 21, 21, 21, 21, 6	3	16.9688		
13. C_{40}	13	3, 9, 18, 21, 24, 24, 15, 6	6	17.1563	17.3672	694.6749
	16	3, 9, 18, 24, 30, 21, 12, 3	6	17.7422		
	18	3, 9, 18, 27, 27, 24, 9, 3	3	17.9766		
	21	3, 9, 18, 24, 27, 21, 15, 3	3	17.6016		
	22	3, 9, 18, 24, 24, 21, 18, 3	3	17.4609		
14. C_{42}	30	3, 9, 18, 24, 27, 27, 9, 3	3	17.6953	17.5212	735.8904
	40	3, 9, 18, 18, 18, 18, 18, 18	1	16.3594		
	1	3, 9, 18, 24, 30, 24, 9, 3	10	17.7891		
	6	3, 9, 18, 24, 24, 24, 15, 3	10	17.5078		
	7	3, 9, 18, 21, 24, 21, 15, 9	20	17.0859		
15. C_{44}	1	3, 9, 18, 24, 27, 24, 15, 6	6	17.7188	17.6271	775.5924
	2	3, 9, 18, 24, 27, 21, 18, 6	6	17.6719		
	3	3, 9, 18, 24, 27, 24, 15, 6	6	17.7188		
	4	3, 9, 18, 21, 27, 24, 15, 9	6	17.3672		
	20	3, 9, 18, 21, 24, 24, 18, 9	6	17.2266		
16. C_{44}	21	3, 9, 18, 21, 27, 24, 15, 9	6	17.3672	17.6511	776.6484
	22	3, 9, 18, 24, 24, 24, 18, 6	6	17.5781		
	1	3, 9, 18, 18, 27, 27, 18, 9, 3	4	17.1445		
	2	3, 9, 18, 21, 27, 27, 18, 9	12	17.5078		
	3	3, 9, 18, 24, 27, 24, 15, 12	12	17.7656		
17. C_{46}	5	3, 9, 18, 24, 24, 24, 21, 9	12	17.6484	17.7702	817.4292
	22	3, 9, 18, 27, 27, 18, 18, 9, 3	4	17.9883		
	1	3, 9, 18, 18, 27, 27, 18, 9, 3	2	17.1445		
	2	3, 9, 18, 21, 27, 24, 15, 15	6	17.4141		
	5	3, 9, 18, 24, 24, 24, 21, 9	12	17.6484		
18. C_{46}	11	3, 9, 18, 24, 27, 24, 18, 9	12	17.7891	17.7702	817.4292
	17	3, 9, 18, 21, 30, 27, 18, 6	6	17.6719		
	20	3, 9, 18, 24, 27, 24, 15, 12	6	17.7656		
	1	3, 9, 18, 24, 27, 27, 18, 12	3	17.9063		
	2	3, 9, 18, 24, 27, 24, 21, 9, 3	3	17.8477		
19. C_{46}	4	3, 9, 18, 21, 27, 27, 18, 15	3	17.5545	17.7702	817.4292
	5	3, 9, 18, 21, 27, 27, 21, 12	3	17.5781		
	6	3, 9, 18, 24, 30, 24, 18, 9, 3	3	17.9883		

Table 1.7. (Cont.)

Fullerene	Atom	DD sequence	μ	ξ	ξ_{av}	$n \xi_{av}$
18. C ₄₈	7	3, 9, 18, 21, 24, 27, 24, 12	3	17.4375	17.7651	852.7248
	8	3, 9, 18, 21, 27, 27, 21, 9, 3	3	17.5664		
	22	3, 9, 18, 24, 30, 24, 15, 15	3	17.9766		
	23	3, 9, 18, 27, 27, 24, 18, 12	3	18.1875		
	24	3, 9, 18, 24, 27, 24, 21, 9, 3	3	17.8477		
	25	3, 9, 18, 21, 24, 24, 21, 18	3	17.3438		
	34	3, 9, 18, 21, 27, 27, 18, 15	3	17.5545		
	35	3, 9, 18, 24, 24, 24, 24, 9, 3	3	17.7070		
	36	3, 9, 18, 24, 27, 27, 18, 12	3	17.9063		
	40	3, 9, 18, 24, 30, 24, 18, 9, 3	3	17.9883		
	42	3, 9, 18, 27, 27, 27, 18, 9	1	18.2578		
	1	3, 9, 18, 24, 27, 24, 24, 12, 3	6	17.9180		
	2	3, 9, 18, 24, 27, 27, 21, 12, 3	6	17.9648		
	7	3, 9, 18, 21, 27, 27, 21, 18	6	17.6250		
	9	3, 9, 18, 24, 27, 27, 24, 9, 3	6	17.9883		
19. C ₅₀	13	3, 9, 18, 27, 30, 24, 18, 15	6	17.3984		
	14	3, 9, 18, 24, 30, 27, 18, 12, 3	6	18.1055		
	15	3, 9, 18, 21, 27, 27, 24, 12, 3	6	17.6367		
	16	3, 9, 18, 21, 24, 27, 24, 18	6	17.4844		
	1	3, 9, 18, 24, 30, 27, 21, 12, 6	10	18.1641	18.0211	901.0550
	6	3, 9, 18, 24, 27, 27, 24, 18	10	18.0469		
20. C ₆₀	11	3, 9, 18, 24, 27, 27, 24, 15, 3	20	18.0352		
	26	3, 9, 18, 21, 30, 27, 21, 18, 3	10	17.8242		
	1	3, 9, 18, 24, 30, 30, 30, 24, 9, 3	60	18.5098	18.5098	1110.5880

distance degree sequences, which are defined as the cardinality of each equivalence class. Next are listed augmented valences for atoms of different equivalence class values (ξ), which represent an index of local atomic complexity. They can be readily computed by multiplying (scalar or inner multiplications of vectors) the corresponding DDS vectors with a vector the components of which are given by inverse powers of 2, that is the vector: (1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, ...). In the following column we give the average value of the local atomic complexity index (ξ_{av}). Finally in the last column, shown as ($n\xi_{av}$), is listed the molecular complexity index obtained by summing the contributions of complexities of all atoms in a molecule, or alternatively obtained by multiplying the average local atomic complexity (ξ_{av}) with the number n of carbon atoms.

There are several interesting features revealed by Table 1.7 worth noting. First, observe that all DDSs start with 3, 9, 18, ..., and this reflects the fact that the smallest rings in fullerenes are five- and six-membered rings. More interesting is the observation that within a single fullerene there

are symmetry non-equivalent carbon atoms with an identical DDS. This is already the case with C_{26} , where two carbon atoms which belong to different orbits have the same DD sequence 3, 9, 18, 21, 18, 9 (these are carbon atoms 5 and 11 in Figure 5 of ref. 52). Additional such instances we find in C_{32} , C_{34} , C_{36} , C_{38} , C_{40} (# 12). In the case of C_{42} there are two pairs of identical DDS values belonging to different symmetry-nonequivalent carbon atoms, and in the least symmetrical fullerene C_{46} we find even four such distance degree sequences. Apparently it is not so uncommon to have non-equivalent carbon atoms in smaller fullerenes with the same DDS. This need not be surprising because DD sequences only record the total sum for valence of neighboring carbon atoms at increasing distance and do not record distributions of such atoms. Thus whenever we have the same number of atoms having the same degree at all distances from a carbon atom but in different distributions, the local atomic complexity will turn out to be the same.

A close look at the two last columns of in Table 1.7, both of which represent a global molecular index of complexity, and which only differ in the scaling factor, is of some interest. That the global complexity increases with the number of carbon atoms is to be expected. However, as we see from Table 1.7 also the average local complexity (ξ_{av}) increases with the increase of the fullerene size. Among the twenty fullerenes of Table 1.7 we have three C_{40} (# 11 – 13) and two C_{44} fullerenes (# 15, 16). Two of the C_{40} fullerenes (# 11 and 12) have the same local and global complexity, as they also have identical average DD sequences: 3.00, 9.00, 18.00, 22.50, 25.20, 22.50, 14.40, 5.40. Thus if one would select instead of $1/2^k$ as the weights for distant neighbors a different scaling, one would nevertheless obtain again the same complexity indices for these two isomers of C_{40} . These two isomers, however, differ considerably in their symmetry properties: one (# 11) has only three classes of carbon atoms (of multiplicity 24, 12 and 4), while the other (# 12) has ten equivalent classes (four of multiplicity 6, five of multiplicity 3 and a single class for itself). Clearly, when symmetry properties of fullerene graphs are to be taken in consideration, we may expect (as will be seen later) that these two cases will show considerable difference in their symmetry-related relative complexities.

Another observation is that whereas for each of the fullerenes # 3, 6, 8, 12, 14, and 17 the DDS and ξ values present degenerate (i. e. identical) pairs or triplets, this is not the case of fullerenes # 7 and 9, which have pairs of degenerate DDS, but these correspond to different ξ values.

7. Comparison of Local Atomic Environments

It may be of interest to compare different carbon environments in the same and in different fullerenes. To do this we should compare DD sequences, which characterize the local complexity of fullerenes. The comparison of sequences, however, need not be as straightforward as is the case of comparison of numbers, because in the case of sequences it is not uncommon to come across pair of sequences that cannot be compared. Consider for illustration the three sequences at the top of Table 1.7 the first of which belongs to C_{20} and the other two to C_{24} , which we will designate as A, B and C, respectively:

$$A = 3, 9, 18, 18, 9, 3 \quad (7.1)$$

$$B = 3, 9, 18, 21, 15, 6 \quad (7.2)$$

$$C = 3, 9, 18, 18, 15, 9 \quad (7.3)$$

It is not difficult to see that both sequences B and C dominate the sequence A, because when we compare member by member in the sequences we see that the corresponding member in sequences B and C are equal to, or larger than the corresponding member in sequence A. Because the entries in the sequences A, B, C represent augmented valences we may conclude that local environments in C_{24} are more complex than the local environment in C_{20} . Observe that this conclusion (which may have been expected in view of the relative size differences between the two fullerenes under consideration) follows without assuming any numerical values for various contributions of distant neighbors to the complexity index.

Of more interest, however, is to compare the sequences B and C within the same fullerene, but here we have a problem: In sequence B we see that b_4 (the fourth entry of the sequence) dominates c_4 , but at the end the opposite happens, entry c_6 dominates entry b_6 . So without additional considerations that would indicate not only which entry in a sequence is more important but also numerically, how much more important, we could not decide on the relative complexity of the two local environments because the corresponding sequences are not comparable. Situations like this are not only common but occur in certain problems so often that no useful comparison is possible. However, about a hundred years ago the Scottish mathematician Muirhead [53] studied the comparability of sequences and introduced an

approach that allows comparison of sequences which otherwise would not be comparable. According to Muirhead, if two sequences:

$$A = a_1, a_2, a_3, \dots a_k \quad (7.4)$$

$$B = b_1, b_2, b_3, \dots b \quad (7.5)$$

cannot be compared, as it happens with the two sequences of fullerene C_{24} , one then constructs the corresponding sequences of partial sums, S_A and S_B , defined as follows:

$$S_A = a_1, a_1 + a_2, a_1 + a_2 + a_3, \dots a_1 + a_2 + a_3 + \dots + a_k \quad (2.6)$$

$$S_B = b_1, b_1 + b_2, b_1 + b_2 + b_3, \dots b_1 + b_2 + b_3 + \dots + b_k \quad (2.7)$$

As we see, at each step instead of considering the initial members of the sequence one considers the corresponding partial sums (the sum of all previous entries in the sequence). Now, instead of comparing the sequences A and B , one compares the novel augmented sequences S_A and S_B . This leads to the set of inequalities, ending with the last relation as the equality:

$$\begin{aligned} a_1 &\geq b_1 \\ a_1 + a_2 &\geq b_1 + b_2 \\ a_1 + a_2 + a_3 &\geq b_1 + b_2 + b_3 \\ &\dots\dots\dots \\ a_1 + a_2 + a_3 + \dots + a_k &= b_1 + b_2 + b_3 + \dots + b_k \end{aligned} \quad (2.8)$$

If all the above relations are satisfied, then we can say that sequence A dominates B , or that sequence B is dominated by A . However, if at least one of the inequalities is not satisfied we say that A and B sequences are not comparable.

If we now return to the two sequences of C_{24} we have instead of:

3, 9, 18, 21, 15, 6

3, 9, 18, 18, 15, 9

the sequences of partial sums:

3, 12, 30, 51, 66, 72

3, 12, 30, 48, 63, 72

and thus clearly the upper sequence dominates the lower sequence, because the corresponding entries in the upper sequence are always either equal to, or larger than the corresponding entries in the lower sequence.

Hence, we can conclude that the atomic environment described by the first sequence is more complex than that of the second type of carbon atom – and again we arrived at this conclusion without using numerical parameters for characterizing contributions of atoms at different distance from the atom under consideration. It is this non-numerical aspect of comparability of sequences, which leads to partial order of different environments, which makes attractive the use of partial order.

In the case of C_{26} we similarly find that SDD 3, 9, 18, 21, 18, 9 dominates the sequence 3, 9, 18, 18, 18, 9, 3, which in turns dominates the sequence 3, 9, 18, 18, 15, 15. But in some fullerenes we have neighborhoods of carbon atoms (which is what DDS represents) that are not comparable. Thus in C_{34} the sequence 3, 9, 18, 21, 24, 15, 9, 3 and the sequence 3, 9, 18, 18, 21, 21, 12 are not comparable, because the corresponding partial sums: 3, 12, 30, 51, 75, 90, 99, 102 and 3, 12, 30, 48, 69, 90, 102, 102 do not satisfy all the inequalities of Muirhead. Observe that after 51, 75, and 90 in the first sequence which dominate the corresponding entries 48, 69, and 90 we have 99 in the first sequence, which does not dominate 102 in the second sequence.

For introductory papers on partial order in chemistry we would like to direct readers to the special issue of communications in mathematical and in computer chemistry *MATCH Communications in Mathematical and in Computer Chemistry* [54], and particularly to articles by Klein (Prolegomenon on Partial Ordering in Chemistry) [55], by Bertz and Zamfirescu on new complexity indices [56], by one of the present authors in collaboration with Vračko, Novič and Basak on ordering of folded structures [57], by El-Basil on partial order among benzenoids [58], and an article of Klein and Bytautas on directed reaction graphs as partial order [59]. Additional illustrations of the use of partial order in chemistry can be found in the book *Order Theoretical Tools in Environmental Sciences* edited by Vogt and Welz [60].

In Table 1.8 we show the partial sums for the ten non-equivalence classes of one of the isomers of C_{40} (# 12) shown in Fig. 1.7. In Fig. 1.8 we have displayed the Hasse diagram induced by the set of ten sequences. Because the second and the fourth sequence are identical we have in Fig. 1.8 only nine partial sum sequences. Because all sequences start with 3, 12, 30 and end with 120 we have suppressed these constant entries in Fig. 1.8 as they do not influence the partial order. In addition, in Table 1.8 we have emphasized the four members of the partial sums shown in Fig. 1.8, which are critical for partial order. As we move in Fig. 1.8 from the top to the

Table 1.8. Sequences of neighbor valence sums and sequences of partial sums for carbon atoms for one of the C_{40} fullerenes (# 12)

Atom	Neighbor valence sum	Partial sums	
1	3, 9, 18, 21, 24, 21, 15, 9	3, 12, 30, 51, 75, 96, 111, 120	17.0859
6	3, 9, 18, 21, 24, 24, 15, 6	3, 12, 30, 51, 75, 99, 114, 120	17.1563
10	3, 9, 18, 21, 21, 21, 21, 6	3, 12, 30, 51, 72, 93, 114, 120	16.9688
13	3, 9, 18, 21, 24, 24, 15, 6	3, 12, 30, 51, 75, 99, 114, 120	17.1563
16	3, 9, 18, 24, 30, 21, 12, 3	3, 12, 30, 54, 84, 105, 117, 120	17.7422
18	3, 9, 18, 27, 27, 24, 9, 3	3, 12, 30, 57, 84, 108, 117, 120	17.9766
21	3, 9, 18, 24, 27, 21, 15, 3	3, 12, 30, 54, 81, 102, 117, 120	17.6016
22	3, 9, 18, 24, 24, 21, 18, 3	3, 12, 30, 54, 78, 99, 117, 120	17.4609
30	3, 9, 18, 24, 27, 27, 9, 3	3, 12, 30, 54, 81, 108, 117, 120	17.6953
40	3, 9, 18, 18, 18, 18, 18, 18	3, 12, 30, 48, 66, 84, 102, 120	16.3594

bottom we are descending from more complex atomic environments to less complex ones for this C_{40} fullerene. At the right-hand side we displayed the computed local complexity indices of Table 1.8 (also shown in Table 1.7) using the same partial order by simply replacing the corresponding sequence entries with carbon atom complexities. Observe that the order induced by the sequences is fully satisfied by the numerical values of the atomic complexity indices.

The importance of the Muirhead comparability inequalities is that they offer additional insights into similarities and differences of local molecular environments [61], which can be extended to local properties of distinct carbon environments in fullerenes. According to Karamata [62], under certain conditions the dominance relationship among sequences can be extended to dominance relations for selected properties of atoms or molecules represented by such sequences, which may be of interest when considering

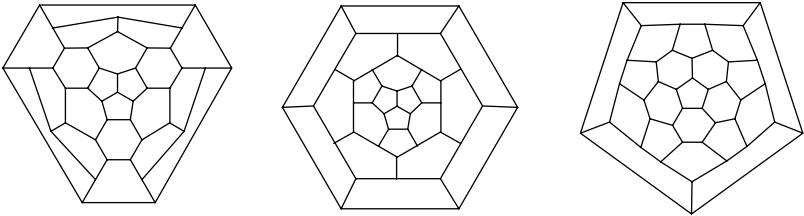


Figure 1.7. Three C_{40} fullerenes discussed in the text (# 11 – 13 in Table 1.7).

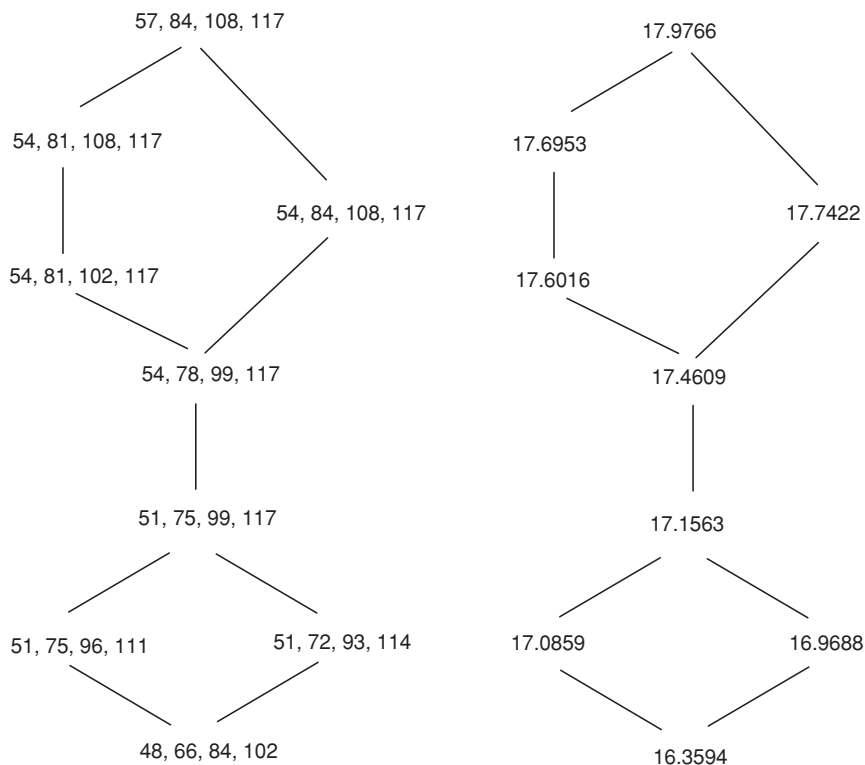


Figure 1.8. Illustration of the partial order on atomic complexity indices for one of C_{40} fullerenes (# 12 in Table 1.7).

local properties of fullerenes. For additional theoretical accounts of the work of Muirhead one could consult more recent books on mathematical inequalities, such as the book *Inequalities* of Beckenbach and Bellman [63].

8. The Role of Symmetry

Most complexity indices of molecular graphs discussed in the literature have not taken into account the presence of molecular symmetry. This in a way may be acceptable when one speaks of graphs, because graphs have no rigid geometry and can be drawn in arbitrary representations, hence also in a very non-symmetrical way. Therefore a molecular index

that is vertex additive need not “know” whether a graph is symmetrical or not. In contrast, complexity measures based on the “information content” of a graph consider partitioning of vertices into equivalence classes and thus automatically relate in some way to molecular symmetry. There is an option when considering the complexity index based on augmented valence [2, 3], which equally applies to other atom-additive type of indices that allow one to select among atoms only one as representative for each class of symmetry-equivalent carbon atoms (as seen in the second column of Table 1.7 and in the first column of Table 1.8). In this way equivalent atoms make a single contribution, as has been the case with the index ξ mentioned in [2, 3].

By restricting summation of atomic contributions only to symmetry-nonequivalent atoms in a molecule we reveal some aspects of molecular symmetry. However, such an approach represents a crude “modification” of the complexity measure to account for the presence of symmetry in structures. The apparent deficiency of such an approach is that it does not consider the size of different equivalence classes. In the construction of complexity indices it seems desirable to take into account the size of each equivalence class. Therefore we propose in this article a novel modification of augmented valence complexity index that takes into account the molecular symmetry by including information on the size of equivalent classes. As will be seen, this novel approach is general and can be applied to other complexity indices, although we will continue to consider the augmented valence as the basis for construction of our complexity index.

We will outline the novel modification of augmented valence complexity index by considering the C_{28} fullerene # 4 of tetrahedral symmetry T_d . In Fig. 1.6 we have already illustrated the Schlegel diagram of C_{28} , which has three distinct DDSs. In Table 1.9 we have summarized information on the three symmetry non-equivalent carbon atoms of C_{28} , which we also used to construct the average DDS (shown in the middle of Table 1.9). Each DD sequence leads to a different local atomic complexity index, which when weighted according to the sizes of equivalence classes (which are 12, 12, 4) gives for the average vertex (or atom) the complexity index $\xi_{av} = 16.1987$. The average vertex complexity index can be obtained also as the dot (or scalar) product of the vector representing the average DDS: {3.0000, 9.0000, 18.0000, 20.5714, 19.2857, 12.8571, 1.2857} and the vector {1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64}.

We will take into account the symmetry of C_{28} by calculating the deviation of each DDS from the average DDS listed above. Here we will

Table 1.9. Construction of the “symmetry factors” illustrated for the C_{28} fullerene

$DDS_1 = 3, 9, 18, 21, 21, 9, 3$	$\xi = 16.2656$	multiplicity 12
$DDS_2 = 3, 9, 18, 21, 18, 15$	$\xi = 16.2188$	multiplicity 12
$DDS_3 = 3, 9, 18, 18, 18, 18, 18$	$\xi = 15.9375$	multiplicity 4
Average weighted DDS: 3.0000, 9.0000, 18.0000, 20.5714, 19.2857, 12.8571, 1.2857	$\xi = 16.1987$	
Deviations from the average: $DDS_1 \rightarrow 0.0669 $ $DDS_2 \rightarrow 0.0201 $ $DDS_3 \rightarrow 0.2612 $		
Symmetry factor $x = 1 + 12 (0.06698) + 12 (0.0201) + 4 (0.2612) = 3.0888$		
$\mu\xi = 3.0888 \times 16.1987 = 50.0345$		

simply take the absolute difference between the corresponding entries in the average sequence and individual DD sequences. This gives the values: 0.0669, 0.0201, and 0.2612 for the three equivalence classes of multiplicity 12, 12, and 4, respectively. These derived deviation values are then used to construct a “symmetry factor” of the form $(1 + x)$ to be used indirectly to account for some aspect of molecular symmetry. The value x is given as the product of the above deviation values and the corresponding multiplicities. Clearly in highly symmetric cases when all carbon atoms belong to the same equivalence class the “symmetry factor” reduces to 1. Among the twenty fullerenes considered here, this is the case only for the smallest C_{20} fullerene and the largest C_{60} buckminsterfullerene, both of which are vertex-transitive and have icosahedral symmetry. In such cases therefore the carbon atom complexity or alternatively the atomic complexity multiplied by the size of fullerene gives the measure of fullerene complexity. To obtain a “symmetry-corrected” complexity index for other structures showing different levels of symmetry, we multiply the average vertex complexity index by the symmetry factor. We adopted provisionally the symbol SC for the “symmetry-corrected” complexity index.

In Table 1.10 we have listed the average atomic values, the symmetry factors and the complexity index SC that accounts for molecular symmetry (which is shown in Table 1.10 in boldface). For comparison we have listed in the last column of Table 1.10 the complexity index ξ (mentioned before),

Table 1.10. Average distance degree sequences, mean atomic complexity (AC), symmetry factor x , symmetry-weighted complexity ($SC = \mu AC$, where μ is the multiplicity) and crude symmetry based complexity (CS) for the fullerenes # 1 through 20.

#	Average distance degree sequence	Mean AC	Factor x	SC	CS
1	3.00, 9.00, 18.00, 18.00, 9.00, 3.00	14.91	1.0000	14.91	14.91
2	3.00, 9.00, 18.00, 19.50, 15.00, 7.50	15.61	4.3744	68.28	31.22
3	3.00, 9.00, 18.00, 20.08, 17.31, 10.38, 0.23	15.92	5.0244	79.99	63.45
4	3.00, 9.00, 18.00, 20.57, 19.29, 12.86, 1.29	16.20	3.0888	50.03	48.42
5	3.00, 9.00, 18.00, 21.00, 20.00, 15.00, 4.00	16.41	8.4990	139.44	49.22
6	3.00, 9.00, 18.00, 20.81, 21.94, 15.00, 4.00	16.04	19.0016	304.87	99.30
7	3.00, 9.00, 18.00, 21.71, 22.76, 18.00, 9.00, 0.53	16.81	9.3382	157.02	135.14
8	3.00, 9.00, 18.00, 22.00, 24.00, 20.00, 9.00, 3.00	17.04	5.5012	93.74	51.12
9	3.00, 9.00, 18.00, 22.26, 23.21, 16.64, 12.32, 5.68, 0.16	17.11	21.3160	364.66	84.83
10	3.00, 9.00, 18.00, 22.26, 24.63, 21.32, 12.32, 3.47	17.21	7.3287	126.11	172.34
11	3.00, 9.00, 18.00, 22.50, 25.20, 22.50, 14.40, 5.40	17.36	4.1048	71.25	51.80
12	3.00, 9.00, 18.00, 22.50, 25.20, 22.50, 14.40, 5.40	17.36	13.4308	233.13	173.20
13	3.00, 9.00, 18.00, 22.50, 25.50, 22.50, 13.50, 6.00	17.37	12.251	212.77	52.24
14	3.00, 9.00, 18.00, 22.71, 26.14, 23.57, 16.29, 7.29	17.52	8.2324	144.24	245.30
15	3.00, 9.00, 18.00, 22.91, 26.18, 24.55, 18.00, 9.82, 0.55	17.63	7.7244	136.16	281.95
16	3.00, 9.00, 18.00, 22.30, 26.59, 24.55, 18.00, 9.82, 0.14	17.65	5.3954	95.23	193.10
17	3.00, 9.00, 18.00, 23.09, 27.00, 25.43, 19.57, 10.57, 1.17	17.77	10.8977	196.66	284.65
18	3.00, 9.00, 18.00, 23.25, 27.38, 26.25, 21.75, 13.50, 1.88	17.89	11.9926	241.55	286.24
19	3.00, 9.00, 18.00, 23.04, 28.20, 27.00, 23.22, 15.96, 3.00	18.02	4.9390	89.01	126.32
20	3.0, 9.0, 18.0, 24.0, 30.0, 30.0, 30.0, 24.0, 9.0, 3.0	18.51	1.0000	18.51	18.51

for which here we use the label CS (crude symmetry). This index is obtained by summing only contributions from the symmetry non-equivalent vertices (those listed in Table 1.7). It is of some interest to make a comparison between the two different measures of complexity, both taking into account indirectly some aspects of graph symmetry. The crude index CS only acknowledges the presence of each of the equivalence classes, while the refined index SC takes into account also the multiplicity (the cardinality of the equivalence classes). In the case of C_{20} and C_{60} the two measures of complexity coincide, because there is only a single equivalence class in both cases, but in contrast to some other symmetry-related indices which do not differentiate complexities of C_{20} and C_{60} , in our case these two highly symmetrical fullerenes show quite different complexity values: 14.9063 and 18.5098, respectively.

There is some parallelism between the two approaches and the two resulting complexity indices CS and SC, but also the differences can be also appreciable. Consider for example the fullerenes # 7, 10 and 12, all three belonging to the same symmetry point group C_{3v} . Fullerene 7, which is slightly smaller, has eight equivalence classes, whereas fullerene 10 and fullerene 12 have both ten equivalence classes. The “crude” approach for all three fullerenes produces relatively large (and not very different) complexity values: $CS(7) = 135.14$; $CS(10) = 172.34$ and $CS(12) = 173.20$, the last two being larger because one is adding contributions from eight and ten carbon atoms, one for each equivalence class. The corresponding revised complexity values for the same three fullerenes are: $SC(7) = 157.02$, $SC(10) = 126.11$ and $SC(12) = 233.13$, respectively. Although being of the same order of magnitude as the preceding ones, crude complexities show different relative magnitudes. The SC index points to C_{38} to be less complex than the smaller C_{34} , while both being significantly less complex than C_{40} of the same symmetry. Superficially one may think it counterintuitive that C_{38} is less complex than C_{34} , but one should recall that we are considering complexities of symmetrical objects and not just complexity of fullerene graphs.

As we can see from Table 1.10, the average CS increases rather slowly as the fullerene size increases. The overall $n \xi_{av}$ increases more dramatically with the size of the fullerene, as can be seen from the last column of Table 1.7. The symmetry-related SC however shows neither a simple regularity in terms of fullerene size (as one would expect), nor does it show any simple regularity within the same point group, or between groups and subgroups. For example, in the case of fullerenes 3 (C_{26}) and 9 (C_{38}), both

of which belong to the point group D_{3h} [52], their respective SC indices are 79.99 and 364.66, respectively. Why such a big difference? Although the increase in size may manifest itself in the increasing value of complexity indices, the size cannot be the main reason for the factor of over 4.5 in the relative magnitudes of these two complexity indices, the relative size of fullerenes being less than 1.5 (i.e., 38/26). As we will explain below, the main factor that influences the relative magnitudes of the symmetry-related complexity indices is the departure of the local characterization (here this is the characterization of atomic neighborhood by augmented valence) from the “average” local characterization. In other words, the more alike are the local environments for symmetry non-equivalent carbon atoms, the smaller is the symmetry-related complexity index; vice-versa, the more diverse are local environments, the larger is the symmetry-related complexity index. Thus in the case of fullerene C_{26} the four symmetry non-equivalent carbon atoms deviate from the average value of 15.92 by 0.21; 0.26; 0.11 and 0.11, which are all of relatively small magnitude and when combined (by multiplying by the respective multiplicities 2, 6, 12 and 6 respectively) they yield the factor $x = 4.0244$, which gives for the symmetry factor $(1 + x) = 5.0244$. In contrast, in the case of fullerene C_{38} (# 9), the five symmetry non-equivalent carbon atoms deviate from the average value of 17.11 by 0.81; 0.82; 0.30; 0.47 and 0.75. These deviations are visibly larger, and when combined (by multiplying by the respective multiplicities 2, 6, 12, 12 and 6 respectively) they give a symmetry factor $x = 20.3160$, resulting in the relatively large value for SC of 364.66.

9. Concluding Remarks on the Complexity of Fullerenes

The complexity index CS listed in the last column of Table 1.10 varies from the smallest values of 14.9063 and 18.5098 (belonging to dodecahedral C_{20} and buckminsterfullerene C_{60} , respectively) to the largest values 284.6484 (belonging to C_{46} , which is the least symmetrical fullerene among the twenty fullerenes under consideration) and 286.24 belonging to fullerene C_{48} . If we weight the carbon atom complexities ξ_{av} by their respective multiplicity, which is tantamount to adding the carbon atom complexity for all atoms in a fullerene, we obtain the values $n \xi_{av}$ shown in the last column of Table 1.7, the strong dependence of which on the size of fullerene is reflected in the range of values that vary from close

to 300 (for C_{20}) to over 1000 (for C_{60}). Because all the atomic values are approximately constant (they vary only between 14.90 and 18.51), there is little novelty in what the $n \xi_{av}$ index reveals. However, there is one remarkable coincidence: two isomers of three C_{40} fullerenes, belonging to different symmetry point groups show an almost identical $n \xi_{av}$ value of 694.31. The more symmetrical C_{40} (# 11), having only three equivalence classes, belongs to the tetrahedral point group T_d , while the less symmetrical isomer # 12, having ten equivalence classes, belongs to a subgroup of T_d , namely the point group C_{3v} . When considering various structural invariants and particularly focusing on isomeric structures one can always expect coincidental values for graph invariants. Without further study it is difficult to speculate whether such coincidences are frequent or not for any particular graph invariant. If they happen too often, this would indicate a potential “weakness” of the particular structural invariant for characterizing various molecular properties, including complexity, as it would imply that too many isomeric species would have the same complexity – which need not be a contradiction, but only makes the particular approach of lesser discriminating power.

By analogy with various molecular descriptors (topological indices), in the case of high degeneracy of molecular descriptors one may consider as a remedy a combination of indices as supplementary descriptors. Thus, it seems prudent to extend the same analogy to molecular complexity and consider a set of complexity indices, or even better an ordered set or sequence of indices, as complexity descriptor. After all, complexity, just as several other molecular attributes, e. g., shape, molecular surface, even molecular size, may exhibit a variation of possible appearances that a single descriptor cannot characterize adequately. In this respect we are suggesting that the average distance degree sequences, listed in Table 1.10 for the twenty smaller fullerenes under consideration, be viewed as components of *complexity vectors*. The sum of the components of these complexity vectors gives $3N$, where N is the number of carbon atoms in fullerene and 3 comes because the degree of carbon atoms in fullerenes is 3.

The idea of representing molecular complexity by vectors offers the possibility to discriminate among structures characterized by the same complexity, as has been the case of two “degenerate” C_{40} fullerenes. However, in this case we find out that not only the two fullerenes have the same complexity index but are also represented by the same “average DDS”, as seen in Table 1.10 for fullerenes # 11 and 12:

3.00, 9.00, 18.00, 22.50, 25.50, 22.50, 12.00, 5.40

Although the introduction of the particular DDS vectors did not resolve the “degeneracy” between these two C_{40} fullerenes, it offers a possibility of ordering not only carbon environments within a single fullerene that has been mentioned previously, but also fullerene structures, and fullerene isomers in particular. For example the other C_{40} fullerene (# 13) has the DD sequence

3.00, 9.00, 18.00, 22.50, 25.50, 22.50, 13.50, 6.00

Using the Muirhead approach for comparison of sequences, we can immediately see that fullerenes # 11 or # 12 dominate fullerene # 13.

10. On the Complexity of Carbon Nanotubes

We will end this discussion by considering some structural aspects that may offer measures of complexity for nanotubes. Because nanotubes typically have a very large number of carbon atoms, it no longer seems practical to take a close look at the neighborhoods of individual carbon atoms, and thus other structural features have to be considered.

10.1. Introductory remarks

Among all the elements of the Periodic Table, carbon has the astounding property of being able to form stable chains and rings of any magnitude. Only three other elements have this property to a degree: boron, silicon and sulfur. Boron [65] is the only one that can form π -bonds with significant strength similarly to carbon, and the existence of borazine, borazaro and boroxaro compounds is proof for this similarity. However, its electron deficit relatively to carbon makes boron a unique element in its propensity of forming two-electron–three-center bonds leading to various types of clusters and to the interesting boranes and carboranes. Silicon, although tetravalent like carbon, lacks its ability of forming stable π -bonds and this fact makes CO_2 and SiO_2 so different. Elemental silicon is isostructural with diamond (and does not form any graphite analog), but silanes are very different from alkanes because they ignite spontaneously in air and are rapidly hydrolyzed by aqueous bases, although they are stable to acids and water [66]. Divalent sulfur can yield various allotropes (but no other compound), and so can silicon, which in addition does give rise to many chemical compounds, but because its sp^2 -hybridized state has

distinctly lower bond energy, it does not rival carbon with its many aromatic derivatives that make organic chemistry much richer than any other class of compounds. The “vegetal and animal kingdoms,” i. e. the constituents of all living organisms, represent but a tiny fraction of these carbon containing compounds.

In addition to the infinite possibilities of organic chemistry, based on carbon as a single chain-forming element, one may conceive two-element alternative chemistries. One of these is based on silicon and oxygen; indeed silicates offer also an infinite number of possible compounds, and constitute the major “mineral kingdom”. Another two-element pair is formed by boron and nitrogen, and indeed boron nitride (BN) also gives rise to two allotropes with similar structures with diamond and graphite [65]. The hexagonal BN (a white non-conducting solid) forms planar sheets that are not staggered like graphite, but have alternating B/N atoms in eclipsed positions of successive layers; the cubic BN analog of diamond is almost as hard, and may replace it in tools. More information on aromatic heterocycles involving B, Si, S and other elements can be found in a recently published review [67].

The traditional allotropes of carbon are the two 3-dimensional diamond lattices with sp^3 -hybridization (cubic diamond, the hardest solid, with ABAB layers, and hexagonal diamond, lonsdaleite, having ABCABC layers with some eclipsed bonds), and the two forms of planar graphite with sp^2 -hybridization (hexagonal graphite with ABAB layers and rhombohedral graphite with ABCABC layers). The first investigation of alternative possibilities of infinite lattices in addition to the above forms was published in 1968 [68]. Later, various other alternatives were investigated [69-71] and the possibility of a regular combination of sp^2 - and sp^3 -hybridized carbon atoms was also considered [72]. The known thermally-favored transition between diamond and the thermodynamically favored graphite at normal pressure must involve gradual interconversion of the two nets, and some theoretical possibilities were examined for this transition, as well as for the reverse process (diamond synthesis at high pressure and temperature). Both these processes must also involve combinations of sp^2 - and sp^3 -hybridized carbon atoms, but without any regularity [73-75].

On rolling a graphene sheet one may convert it into “buckycones” (predicted [76] before their experimental detection [77], and briefly discussed later) or carbon nanotubes. Before discussing nanotubes, however, one has to emphasize that like all synthetic polymers and many natural

macromolecules such as polysaccharides (starch, glycogen, and cellulose), nanotubes are not substances, strictly speaking. Indeed, their molecules are similar (but not identical), and differ in the polymerization degree and molecular weight; their resulting polydispersity may be broad or narrow. Only two classes of natural polymers are monodisperse: proteins and polynucleotides. So far, it has not yet become possible to obtain monodisperse carbon nanotubes.

It is well known since Euler's time that in order to obtain a polyhedral molecule formed from hexagons and pentagons, there must be exactly twelve pentagons. The presence of odd-membered rings explains why BN analogues of fullerenes are not stable. By contrast BN analogues of buckycones and carbon nanotubes are possible, but they are not yet known, therefore in the following when we will discuss nanotubes it will be understood that they are carbon nanotubes.

In an interesting survey of the number of publications since the discovery of fullerenes [78] (with their large-scale synthesis five years later [79]) and carbon nanotubes [80, 81], it was found that the number of publications about fullerenes has reached a plateau, whereas those about nanotubes continue to increase exponentially [82].

10.2. Helicity of nanotubes

On rolling a long strip of a planar graphene sheet into a nanotube, the result depends on how the ends of the strip become connected. We shall adopt the notation introduced by Klein et al. [83] that defined the geometry of the buckytube in terms of two parameters: the *twist* (t_+) and the *counter twist* (t_-) with the conditions that t_+ must be an integer higher than, or equal to, 3, whereas t_- can have any integer value between 0 and t_+ ; in other words, $t_- \leq t_+$.

The simplest way to visualize the result is to express the geometry of the nanotube as a function of the helicity given by the two integer numbers denoted by t_+ and t_- and associated with the dualist graph of the benzenoid rings in the graphene strip (this graph has the centers of hexagons as vertices, and its edges connect vertices corresponding to condensed benzenoid rings). The numbers t_+ and t_- indicate how many linearly condensed hexagons in two directions diverging from one hexagon by 120° in the nanotube are involved in the repeating pattern on the nanotube till they overlap when rolling the strip (Fig. 1.9). The two extremes are nanotubes with

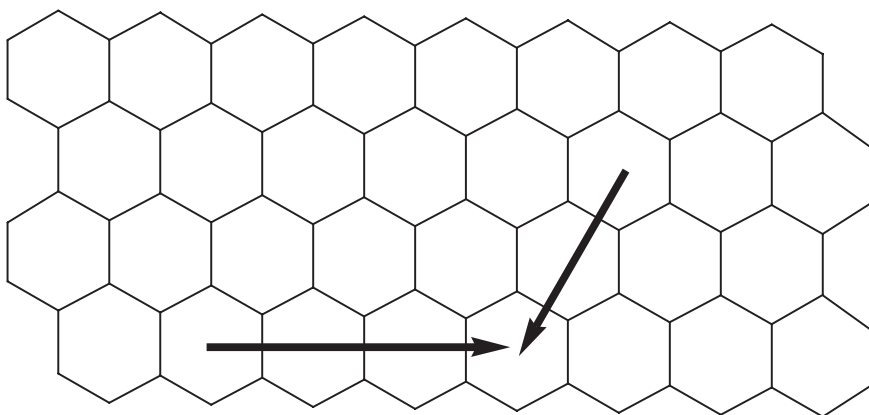


Figure 1.9. A portion of the graphite net with $t_+ = 4$, $t_- = 2$.

$t_- = 0$ and $t_+ = t_-$; they are achiral, and are known as *zigzag* and *armchair* nanotube, respectively. In other words, when the countertwist is zero, the buckytube is achiral and is formed by circles of acenes and is called a “zigzag” nanotube. Its diameter D (in Ångstroms) is approximately $0.78t_+$; when $t_+ = t_-$ the nanotube is also achiral and is called an “armchair” nanotube, and its diameter D (in Å) is $0.78t_+\sqrt{3}$; finally, in the general case with $t_+ \neq t_-$ the nanotube is chiral and has a diameter (in Å) given by Eq. (10.1):

$$D = 0.78 (t_+^2 + t_-^2 + t_+t_-)^{1/2} (\text{in Ångstroms}) \quad (10.1)$$

In Figs. 1.10A to 1.10C we present stereo-views of three examples for the preceding three cases, all with diameters close to 4 Å. The drawings are accompanied by data about the number of benzenoid rings of nanotubes with open ends, their approximate diameters, and strain energies calculated by molecular mechanics (MM2), computed and drawn by means of the CambridgeSoft Chem3D Program.

One can define the helicity H by the following expression:

$$H = \sin \left(1 - \frac{t_-}{t_+} \right) \pi \quad (10.2)$$

For simplicity, we will consider only single-wall nanotubes of various lengths, ignoring any fullerenic caps that are known to exist in experimentally observed nanotubes. Fig. 1.11 illustrates a zigzag nanotube with one

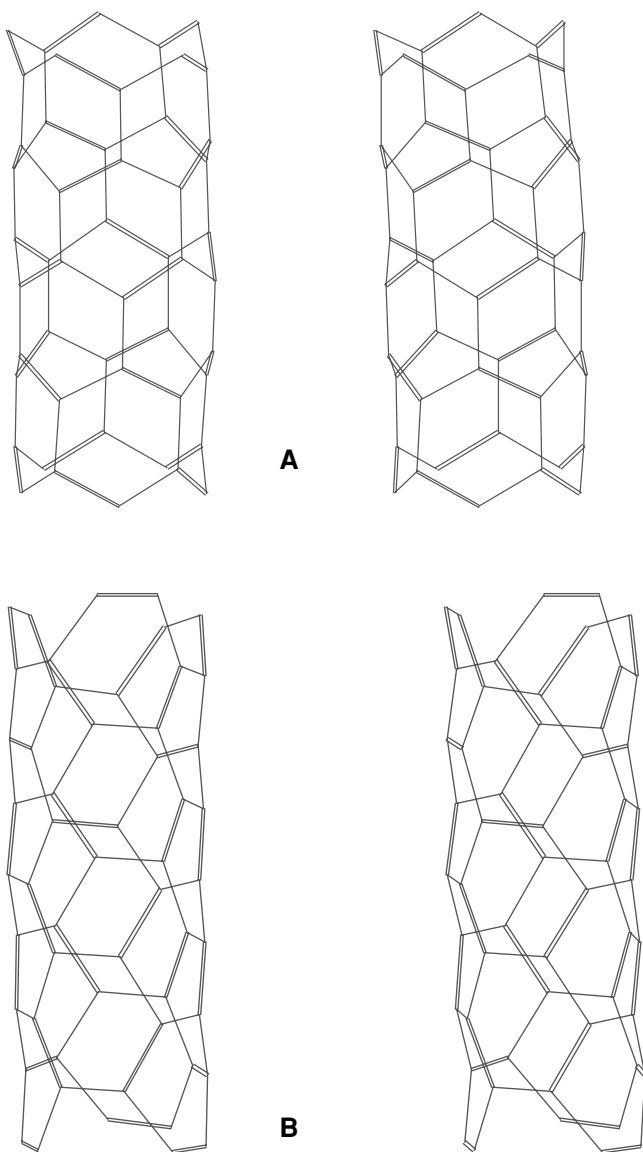


Figure 1.10. (A) Stereo-view of an achiral zigzag nanotube with open ends having $t_+ = 5$, $t_- = 0$, with open ends. It has 20 benzenoid rings, diameter 3.9 Å, and strain energy (MM2) 292.89 kcal/mol (14.64 kcal/mol per hexagon). (B) Stereo-view of an achiral armchair nanotube having $t_+ = t_- = 3$, with open ends. It has 21 benzenoid rings, diameter 4.05 Å, and strain energy 237.03 kcal/mol (11.29 kcal/mol per hexagon).(*cont.*)

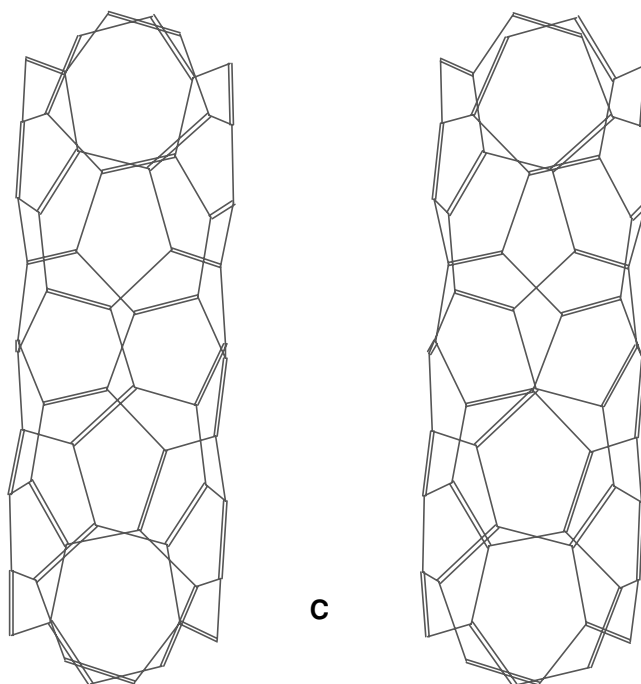


Figure 1.10. (Cont.) (C) Stereo-view of a chiral nanotube having $t_+ = 4$, $t_- = 2$, with open ends. It has 24 benzenoid rings, diameter 4.13 Å, and strain energy (MM2) 245.92 kcal/mol (10.25 kcal/mol per hexagon).

capped end. One may also speculate that heteroatoms might simplify these ends of nanotubes [84, 85]. Also, nanotubes will be considered to be constituted exclusively of benzenoid rings, without any kinks caused by 5- or 7-membered rings.

In principle, as discussed earlier in this review, higher symmetry implies smaller complexity. One may argue therefore that the longer the nanotube, the higher its symmetry. On the other hand, the complexity of molecules increases with its number of atoms. Thus, the lengthening of the nanotube can both increase the complexity by adding carbon atoms, and decrease the complexity by enhancing the symmetry in a more subtle fashion. Therefore we will consider only the complexity derived from the parameters t_+ and t_- , which define both the helicity (Eq. 10.2) and the diameter (Eq. 10.1).

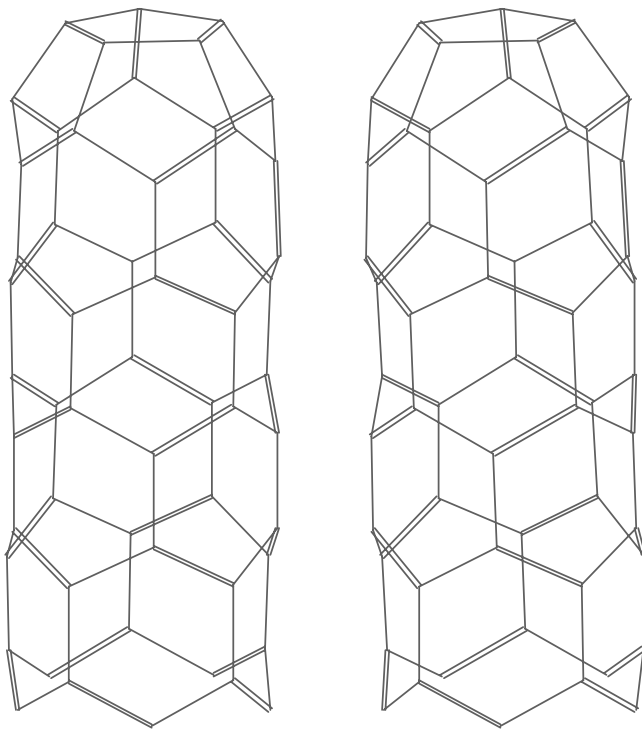


Figure 1.11. Stereo-view of a zigzag nanotube with one capped end, having $t_+ = 5$, $t_- = 0$.

No single parameter seems to be “definitory” for the complexity, for instance by compensating for increased complexity due to a larger diameter by the decreased complexity due to lower or zero helicity.

The complexity of a nanotube will depend on several parameters: length, diameter, helicity. The difference $t_+ - t_-$ that is involved in Eq. (10.2) provides a measure of helicity. The values for the ratio t_-/t_+ range from zero (for zigzag nanotubes) through 0.5 (for the skew nanotube shown in Fig. 10C) to 1 for $t_- = t_+$ (for armchair nanotubes), and consequently the values for the helicity H range from zero (for zigzag and armchair nanotubes) to 1 for $t_-/t_+ = 0.5$. It seems that the highest possible complexity due to helicity as indicated by Eq. (10.1) should be assigned to the case when $t_-/t_+ = 0.5$ (Fig. 10C), at equal “informational distance” from the two achiral nanotubes illustrated in Figs. 10A and 10B. The second variable

determining the complexity may be considered to be the nanotube diameter, given by Eq. (10.1), so that the two parameters (t_+ and t_-) are sufficient to define both the diameter and the helicity. The same two parameters (actually the difference $t_+ - t_-$, namely if it is or is not divisible by 3) determine the HOMO–LUMO gap in nanotubes; this gap in turn influences their electrical conductivity [86]. A discussion on the aromaticity of armchair nanotubes may be found in [87].

M. V. Diudea, T. S. Balaban, E. C. Kirby and A. Graovac have recently published a study on how to derive single-walled nanotubes from cylindrical surfaces generated from a square net by deleting bonds in order to change squares into hexagons [88]. Their end result is the prediction of metallic or semiconductor behavior of the nanotube in terms of characteristics of the parent square lattice. This study contains further references to papers by Diudea and Kirby on this topic, including ones that allow coding of the nanotube or carbon torus structure [89–92].

Since the length, the possible capping, bending, and presence of multi-walled nanotubes will further complicate the picture, we prefer to terminate here the discussion of nanotube complexity.

Discussing the complexity of few *other types of carbon nanostructures* has not been included in the present chapter, but leading references will be included in the following. For *carbon tori*, in addition to the literature cited above [89–92], it must be mentioned that they were discussed in the literature [93–95] before being obtained experimentally along with nanotubes and initially called “crop circles” [96]. *Coiled nanotubes* [97–99], *hyperfullerenes* with negative curvature [100], and *Y-junctions* of carbon nanotubes [101] are closely related to nanotubes. However, *graphitic cones* are quite different, and were predicted [76] before they were observed for the first time [77]. Subsequently, they were found to occur naturally [102], and were also obtained synthetically [103]. Some of their topological characteristics have also been discussed [104–105].

Acknowledgements

ATB thanks Professor D. J. Klein for discussions that will form the basis of a future joint paper on the complexity of nanotubes. MR would like to thank Professor Jure Zupan for kind hospitality during his visit to the National Institute of Chemistry, Ljubljana, Slovenia, and XG is grateful to National Science Foundation of China, which in part supported the work on this Project.

References

1. H. Primas, *Chemistry, Quantum Mechanics and Reductionism (Perspectives in Theoretical Chemistry)*, 2nd ed, Springer, Berlin (1983) p. 292.
2. M. Randić and D. Plavšić, *Int. J. Quantum Chem.* **91**, 20-31 (2003).
3. M. Randić and D. Plavšić, *Croat. Chem. Acta* **75**, 107-116 (2002).
4. R. Barone and M. Chanon, *J. Chem. Inf. Comput. Sci.* **41**, 269-272 (2001).
5. D. Bonchev, *J. Chem. Inf. Comput. Sci.* **40**, 934-941 (2000).
6. S. Nikolić, N. Trinajstić and I.M. Tolić, *J. Chem. Inf. Comput. Sci.* **40**, 920-926 (2000); S. Nikolić, I.M. Tolić and N. Trinajstić, *MATCH Commun. Math. Comput. Chem.* **40**, 187-201 (1999).
7. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **40**, 99-106 (2000).
8. S.H. Bertz, *Bull. Math. Biol.* **45**, 849-855 (1983).
9. S.H. Bertz, *J. Am. Chem. Soc.*, **103**, 3599-3601 (1981).
10. S.H. Bertz, in: *Chemical Applications of Topology and Graph Theory*, R B King, (ed.), Elsevier, New York (1983) p. 206-221.
11. H. Wiener, *J. Am. Chem. Soc.* **69**, 17-20 (1947).
12. P.E. John, R.B. Mallion and I. Gutman, *J. Chem. Inf. Comput. Sci.* **38**, 108-112 (1998).
13. I. Gutman, R. B. Mallion and J.W. Essam, *Mol. Phys.* **50**, 859-877 (1983).
14. S. Nikolić, N. Trinajstić, A. Jurić, Z. Mihalić and G. Krilov, *Croat. Chem. Acta* **69**, 883-897 (1996).
15. M. Randić, G. M. Brissey, R. G. Spencer and C. L. Wilkins, *Computers & Chem.* **3**, 5-13 (1979).
16. A. Mowshowitz, *Bull. Math. Biophys.* **30**, 175-204 (1968).
17. C. Shannon and W. Weaver, *Mathematical Theory of Communications*, University of Illinois Press, Urbana, IL (1949).
18. D. Bonchev and O. E. Polansky, in *Graph Theory and Topology in Chemistry*, R. B. King and D. H. Rouvray, (eds.), Elsevier, Amsterdam (1987) p. 126-158.
19. H. Primas, *Chemistry, Quantum Mechanics and Reductionism (Perspectives in Theoretical Chemistry)*, 2nd ed., Springer, Berlin (1983) p. 254.
20. R. Hoffmann, *Am. Sci.* **76**, 182-185 (1988).
21. M. Randić, *Chem. Rev.* **103**, 3449-3605 (2003).
22. M. Randić, *Chem. Phys. Lett.* **38**, 68-70 (1976),
23. M. Randić, *J. Am. Chem. Soc.* **99**, 444-450 (1977).
24. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA (1969).
25. M. Randić, in: *The Encyclopedia of Computational Chemistry*, P. V. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner (eds.), Wiley, Chichester, p. 3018-3032.
26. A. T. Balaban, *Highly Discriminating Distance-Based Topological Index*. *Chem. Phys. Lett.* **80**, 399-404 (1982).
27. J. Devillers and A. T. Balaban (eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, The Netherlands (1999).

28. M. Randić, *J. Am. Chem. Soc.* **97**, 6609-6615 (1975).
29. M. Randić, *J. Mol. Graphics Model.* **20**, 19-35 (2001).
30. H. L. Morgan, *J. Chem. Docum.* **5**, 107-113 (1965).
31. M. Razinger, *Theor. Chim. Acta*, **61**, 581-586 (1982).
32. J. Figueras, *J. Chem. Inf. Comput. Sci.* **33**, 717-718 (1993).
33. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **33**, 683-695 (1993).
34. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **41**, 1457-1462 (2001).
35. F. Ayers, Jr., *Theory and Problems of Matrices*, Schaum, New York (1962) p. 181.
36. D. M. Cvetković and I. Gutman, *Croat. Chem. Acta*, **49**, 115-121 (1977).
37. L. Lovasz and J. Pelikan, *Period. Math. Hung.* **3**, 175-182 (1973).
38. Comment by a referee.
39. E. Ruch and I. Gutman, *J. Comb. Inf. System Sci.* **4**, 285-295 (1979).
40. M. Randić, J. Zupan, M. Novič, B. D. Gute and S. C. Basak, *SAR & QSAR in Environ. Res.* **13**, 689-703 (2002).
41. M. Randić, M. Novič and M. Vračko, *Eigenvalues as Molecular Descriptors*, in: *QSPR/QSAR Studies by Molecular Descriptors*, M. V. Diudea (ed.), Nova Science, Huntington, NY (2002) p. 147-211.
42. D. J. Klein, *J. Chem. Edu.* **69**, 691 (1992).
43. D. J. Klein and D. Babić, *J. Chem. Inf. Comput. Sci.* **37**, 656-671 (1997).
44. M. Randić, *Chem. Rev.*, **103**, 3449-3605 (2003).
45. M. Randić, *Croat. Chem. Acta* **66**, 289-312 (1993).
46. (a) S. Nikolić, N. Trinajstić, I. M. Tolić, G. Rücker and C. Rücker, *On Molecular Complexity Indices*, in: *Complexity in Chemistry*, D. H. Rouvray and D. Bonchev (eds.), Francis & Taylor, London (2003); (b) A. T. Balaban, *A Comparison Between Various Topological Indices, Particularly Index J and Wiener's Index W* in: *Topology in Chemistry: Discrete Mathematics of Molecules*, D. H. Rouvray and R. B. King (eds.) Horwood, Chichester (2002) pp. 89-112; (c) A. T. Balaban, D. Mills, and S. C. Basak, *MATCH Commun. Math. Comput. Chem.* **45**, 5-26 (2002).
47. M. Randić, *J. Math. Chem.* **25**, 345-358 (1998).
48. M. Randić, M. Razinger and D. Plavšić, *MATCH Commun. Math. Comput. Chem.* **35**, 243-259 (1997).
49. M. Randić, N. Basak and D. Plavšić, *Croat. Chem. Acta* **77**, 251-257 (2004).
50. D. Bonchev, *The Problems of Computing Molecular Complexity*, in: *Computational Chemical Graph Theory*, D. H. Rouvray (ed.) Nova, Commack, NY (1990) p. 33-63; D. Bonchev, *Overall Connectivity and Topological Complexity: A New Tool for QSPR/QSAR*, in: *Topological Indices and Related Descriptors in QSAR and QSPR*, J. Devillers and A. T. Balaban (eds.), Gordon and Breach, The Netherlands (1999), p. 361-401; D. Bonchev and W. A. Seitz, *The Concept of Complexity in Chemistry*, in: *Concepts in Chemistry – A Contemporary Challenge*, D. H. Rouvray (ed.), Wiley, New York (1997) p. 353-381; D. Bonchev, *Shannon's Information and Complexity*, in: *Complexity in Chemistry. Introduction and Fundamentals*, D. Bonchev and D. H. Rouvray (eds.), Taylor and Francis, London (2003) p. 157-187; D. Bonchev,

- MATCH Commun. Math. Comput. Chem. **7**, 65-113 (1979); D. Bonchev, Bulg. Chem. Commun. **28**, 567-582 (1995); D. Bonchev, SAR QSAR Environ. Res. **7**, 23-43 (1997); D. Bonchev, Croat. Chem. Acta **77**, 167-173 (2004).
51. M. Randić, Croat. Chem. Acta **74**, 683-705 (2001).
 52. T. Laidboeur, D. Cabrol-Bass and O. Ivanciuc, J. Chem. Inf. Comput. Sci. **36**, 811-821 (1996). For a complete list of fullerenes and their symmetries, see: P. W. Fowler and D. E. Manolopoulos, An Atlas of Fullerenes, Oxford, Clarendon (1995).
 53. R. F. Muirhead, Proc. Edinburgh Math. Soc. **21**, 144-157 (1903).
 54. Special issue of MATCH Commun. Math. Comput. Chem. **42**, 1-290 (2000).
 55. D. J. Klein, MATCH Commun. Math. Comput. Chem. **42**, 7-31 (2000).
 56. S. H. Bertz and C. M. Zamfirescu, MATCH **42**, 39-70 (2000).
 57. M. Randić, M. Noviĉ and M. Vraĉko, MATCH **42**, 39-231 (2000).
 58. S. El-Basil, MATCH Commun. Math. Comput. Chem. **42**, 233-259 (2000).
 59. D. J. Klein and L. Bytautas, MATCH Commun. Math. Comput. Chem. **42**, 261-290 (2000).
 60. K. Vogt and G. Welz, Order Theoretical Tools in Environmental Sciences, Shaker, Aachen (2002).
 61. M. Randić, Chem. Phys. Lett. **55**, 547-551 (1978).
 62. J. Karamata, Publ. Math. Univ. Belgrade **1**, 145 (1932).
 63. E. F. Beckenbach and R. Bellman, Inequalities, Springer, Berlin, 1961.
 64. G. H Hardy, J. E. Littlewood and G. Polya, Inequalities, Cambridge University Press, London (1934).
 65. F. A. Cotton, G. Wilkinson, C. A. Murillo and M. Bochmann, Advanced Inorganic Chemistry, 6th ed., Wiley, New York, pp. 131-157, 168-180.
 66. F. A. Cotton, G. Wilkinson, C. A. Murillo and M. Bochmann, Advanced Inorganic Chemistry, 6th ed., Wiley, New York, pp. 265-280.
 67. A. T. Balaban, D. Oniciu and A. R. Katritzky, Chem. Revs. **104**, 000 (2004).
 68. A. T. Balaban, C. C. Rentea and E. Ciupitu, Rev. Roum. Chim. **13**, 231-247; Erratum, *ibid.*, p. 1233 (1968).
 69. H. Zhu, A. T. Balaban, D. J. Klein, and T. P. Živković, J. Chem. Phys. **101**, 5281-5292 (1994).
 70. A. T. Balaban, Theoretical Investigation of Carbon Nets and Molecules, in: Theoretical Organic Chemistry, C. Parkanyi (ed.), Elsevier, Amsterdam (1998), p. 381-404.
 71. A. T. Balaban, Computers Math. Applic. **17**, 397-416 (1989); Reprinted in: Symmetry II, I Hargittai (ed.), Pergamon Press, Oxford (1989) p. 397-416.
 72. K. M. Merz Jr, R. Hoffmann and A. T. Balaban, J. Am. Chem. Soc. **109**, 6742-6751 (1987).
 73. A. T. Balaban, D. J. Klein and C. A. Folden, Chem. Phys. Lett. **217**, 266-270 (1994).
 74. A. T. Balaban, D. J. Klein, and W. A. Seitz, Intern. J. Quantum Chem. **60**, 1065-1068 (1996).

75. A. T. Balaban and D. J. Klein, *Carbon* **35**, 247-251 (1997).
76. A. T. Balaban, D. J. Klein, and X. Liu, *Carbon* **32**, 357-359 (1994).
77. M. H. Ge and K. Sattler, *Chem. Phys. Lett.* **220**, 3-5 (1964).
78. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *Nature* **318**, 162-163 (1985).
79. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, *Nature* **347**, 354-358 (1990).
80. S. Iijima, *Nature* **354**, 56-58 (1991).
81. T. W. Ebbesen and P. M. Ajayan, *Nature* **358**, 220-222 (1992).
82. R. N. Kostoff, T. Braun, A. Schubert, D. R. Toothman, and J. A. Humenik, *J. Chem. Inf. Comput. Sci.* **40**, 19-39 (2000).
83. D. J. Klein, W. A. Seitz, and T. G. Schmalz, *J. Phys. Chem.* **97**, 1231-1236 (1993).
84. A. T. Balaban, *MATCH Commun. Math. Comput. Chem.* **33**, 25-33 (1996).
85. A. T. Balaban, *Bull. Soc. Chim. Belges* **105**, 383-389 (1996).
86. H. Y. Zhu, D. J. Klein, T. G. Schmalz, *J. Phys. Chem. Solids* **59**, 417-423 (1998); D. J. Klein and L. Bytautas, *J. Phys. Chem. A* **103**, 5196-5210 (1999); O. Ivanciuc, D. J. Klein and L. Bytautas, *Carbon* **40**, 2063-2083 (2002).
87. I. Lukovits, A. Graovac, E. Kálmán, G. Kaptay, P. Nagy, S. Nikolic, J. Sytchev, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **43**, 609-614 (2003).
88. M. V. Diudea, T. S. Balaban, E. C. Kirby, and A. Graovac, *Phys. Chem. Chem. Phys.* **5**, 4210-4214 (2003).
89. M. V. Diudea, M. Stefu, B. Parv, and P. E. John, *Croat. Chem. Acta* **77**, 111-115 (2004); P. E. John and M. V. Diudea, *ibid.* 127-132.
90. M. V. Diudea, *Carbon Nanostruct.* **10**, 273-292, 2002; M. V. Diudea, *Topology of Naphthylenic Tori*, *Phys. Chem. Chem. Phys.* **4**, 4740-4746 (2002).
91. M. Diudea, I. Silaghi-Dumitrescu, and B. Parv, *Internet Electronic J. Mol. Design* **1**, 10-22 (2002).
92. E. C. Kirby, *Recent Work on Toroidal and Other Exotic Fullerene Structures*, in: *From Chemical Topology to Three-Dimensional Geometry*, A. T. Balaban (ed.), Plenum Press, New York (1997) p. 263-296.
93. M. Terrones, W. K. Hsu, J. P. Hare, H. W. Kroto, H. Terrones and D. R. M. Walton, *Phil. Trans. Roy. Soc. London A* **354**, 2025-2054 (1996).
94. S. Ihara, S. Itoh, and J. I. Kitakami, *Phys. Rev. B* **47**, 12 908-12 911 (1993).
95. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes*, Academic Press, San Diego, CA (1996) p. 756.
96. J. Liu, H. Dai, J. H. Hafner, D. T. Colbert, R. E. Smalley, S. J. Tans, and C. Dekker, *Nature* **385**, 780-781 (1997).
97. B. I. Dunlap, *Phys. Rev. B* **46**, 1933-1936 (1992).
98. S. Ihara, S. Itoh, and J. Kitakami, *Phys. Rev. B* **48**, 5643-5647 (1993).
99. L. P. Biro, G. I. Mark, and P. Lambin, *IEEE Trans. Nanotechnol.* **2**, 362-367 (2003).
100. G. E. Scuseria, *Chem. Phys. Lett.* **195**, 534-536 (1992).

101. A. N. Andriotis, M. Menon, D. Srivastava and L. A. Chernozatonskii, *Phys. Rev. Lett.* **87**, 066 802-1 – 066 802-4 (2001).
102. J. A. Jaszczak, G. W. Robinson, S. Dimovski and Y. Gogotsi, *Carbon* **41**, 2085-2092 (2003).
103. G. Zhang, X. Jiang and E. Wang, *Science* **300**, 472-474 (2003).
104. P. E. Lammert and V. H. Crespi, *Phys. Rev. Lett.* **85**, 5190-5193 (2000).
105. M. M. J. Treacy and J. Kilian, *Mat. Res. Soc. Symp. Proc.* **675**, 1-6 (2001).

Chapter 2

COMPLEXITY AND SELF-ORGANIZATION IN BIOLOGICAL DEVELOPMENT AND EVOLUTION

Stuart A. Newman

New York Medical College, Valhalla, NY 10595

Gabor Forgacs

University of Missouri, Columbia, MO 65211

1. Introduction: Complex Chemical Systems in Biological Development and Evolution

The field of developmental biology has as its major concern *embryogenesis*: the generation of fully-formed organisms from a fertilized egg, the *zygote*. Other issues in this field, organ *regeneration* and tissue *repair* in organisms that have already passed through the embryonic stages, have in common with embryogenesis three interrelated phenomena: *cell differentiation*, the production of distinct cell types, *cell pattern formation*, the generation of specific spatial arrangements of cells of different types, and *morphogenesis*, the molding and shaping of tissues [1]. The cells involved in these developmental processes and outcomes generally have the same genetic information encoded in their DNA, the *genome* of the organism, so that the different cell behaviors are largely associated with *differential gene expression*.

Because of the ubiquity and importance of differential gene expression during development, and the fact that each type of organism has its own unique genome, a highly gene-centered view of development prevailed for several decades after the discovery of DNA's capacity to encode information. In particular, development was held to be the unfolding of a "genetic program" specific to each kind of organism and organ, based on differential gene expression. This view became standard despite the fact that no convincing models had ever been presented for how genes or their products (proteins and RNA molecules) could alone build three

dimensional shapes and forms, or even generate populations of cells that utilized the common pool of genetic information in different ways. By “alone” is meant without the help of physics and chemical dynamics, the scientific disciplines traditionally invoked to explain changes in shape, form, and chemical composition in nonliving material systems.

It has long been recognized that biological systems are also physicochemical systems, and that phenomena first identified in the nonliving world can also provide models for biological processes. Indeed, at the level of structure and function of biomolecules and macromolecular assemblages such as membranes, biology has historically drawn on ideas from chemistry and physics. More recently, however, there has been intensified cooperation between biological and physical scientists at the level of complex biological systems, including developmental systems. Applications of results from these fields at the systems level, however, had to wait until it became clear to modern biologists (as it was to many before the “gene revolution”) that organisms are more than programmed expressions of their genes, and that the behavior of systems of many interacting components is neither obvious nor predictable on the basis of the behavior of their parts. These new approaches also could not have occurred until physical scientists began fully engaging with experimentally-derived details of the phenomena in question (a departure from much of earlier “theoretical biology”), and developed theoretical tools and computational power sufficient to model systems of great complexity. This new systems biology has been emerging over the last decade.

Since biology, with the help of chemistry and physics, is increasingly studied at the systems level, there has also been new attention to the origination of such systems. In many instances, for example, it is reasonable to assume that complexity and integration in living organisms has evolved in the context of forms and functions that originally emerged (in evolutionary history) by straightforward physicochemical means. Thus the elaborate system of balanced, antagonistic signaling interactions that keep cell metabolism homeostatic, and embryogenesis on-track, can be seen as the result of accretion, by natural selection, of stabilizing mechanisms for simpler physicochemical generative processes that would otherwise be less reliable.

The richness of the phenomena of development prohibits any comprehensive review of physicochemical approaches to their analysis in the space of a single chapter. But a rough separation can be made between those processes that are largely the province of physics—the folding and stretching

of cell sheets and the molding and separation of cell masses—and those that are the province of chemical dynamics—switching between alternative stationary compositional states leading to cell differentiation, chemical oscillations resulting in segmental organization of tissues, and breaking of symmetry by reaction-diffusion coupling leading to cellular pattern formation. It is the latter group of processes and phenomena that we will consider below. But even in this narrowed set it will be possible to describe only a limited selection of biological examples and proposed explanatory models. Finally, we will describe a hypothesized scenario by which a chemical-dynamical mechanism of development that plausibly originated a class of embryonic patterns early in the history of multicellular life was transformed over the course of evolution into a more reliable genetically-programmed mechanism for producing the same forms.

In many cases, specific molecules that participate in the chemical-dynamic mechanisms we discuss will be indicated, in recognition both of the enormous progress that has been made in recent years in identifying the genes and gene products involved in complex developmental processes, and of the credibility of proposed theoretical frameworks insofar as they deal with characterized, experimentally-accessible components. But readers mainly interested in the formal aspects of the processes under discussion may look past the molecular names with little disadvantage.

2. Dynamic Multistability and Cell Differentiation

The early embryos of multicellular organisms are referred to as *blastulae*. These are typically (but not invariably) hollow clusters of several dozen to several hundred cells. While the zygote, the fertilized egg that gives rise to the blastula, is “totipotent,” that is, it has the potential to give rise to any of the more than 200 specialized cell types (e.g., the various types of muscle, blood, and nerve cells) of the mature human body, by the blastula stage the cells are no longer identical. In most species, the first few cell divisions in the embryo generate cells that are “pluripotent”—capable of giving rise to only a limited range of cell types. These cells, in turn, diversify into cells with progressively limited potency, ultimately generating all the (generally unipotent) specialized cells of the body [1].

The transition from wider to narrower developmental potency is referred to as *determination*. This stage of cell specialization generally occurs with

no overt change in the appearance of cells. Instead, subtle modifications, only discernable at the molecular level, set the altered cells on new and restricted developmental pathways. A later stage of cell specialization, referred to as *differentiation*, results in cells with vastly different appearances and functional modifications—electrically excitable neurons with extended processes up to a meter long, bone and cartilage cells surrounded by solid matrices, red blood cells capable of soaking up and disbursing oxygen, and so forth. Cells have generally become determined by the end of blastula formation, and successive determination increasingly narrows the fates of their progeny as development progresses. When the developing organism requires specific functions to be performed, cells will typically undergo differentiation.

Since each cell of the organism contains an identical set of genes (except for the egg and sperm and their immediate precursors, and some cells of the immune system), a fundamental question of development is how the same genetic instructions can produce different types of cells. This question pertains to both determination and differentiation. Since these two kinds of cell specialization are formally similar and probably employ overlapping set of molecular mechanisms, we will refer to both as “differentiation” in the following discussion, unless confusion would arise. Multicellular organisms solve the problem of specialization by activating only a type-specific subset of genes in each cell type.

The *biochemical state* of a cell can be defined as the list of all the different types of molecules contained within it, along with their concentrations. The dynamical state of a cell, like that of any dynamical system, resides in a multidimensional space, the “state space,” with dimensionality equal to the number of system variables (e.g., chemical components) [2]. During the cell division cycle (i.e., the sequence of changes that produces two cells from one), also called simply the cell cycle, the biochemical state changes periodically with time. (This, of course, assumes that cells are not undergoing differentiation.) If two cells have the same complement of molecules at corresponding stages of the cell cycle, then, they can be considered to be of the same differentiated state. The cell’s biochemical state also has a spatial aspect—the concentration of a given molecule might not be uniform throughout the cell. We will discuss this in Section 4, below. Certain properties of the biochemical state are highly relevant to understanding developmental mechanisms. The state of differentiation of the cell (its *type*) can be identified with the collection of proteins it is capable of making.

Of the estimated 20,000-25,000 human genes [3], a large proportion constitutes the “housekeeping genes,” involved in functions common to all or most cells types [4]. In contrast, a relatively small number of genes—possibly fewer than a thousand—specify the type of determined or differentiated cells; these genes are *developmentally regulated* (i.e., turned on and off in an embryonic stage- and embryonic position-dependent fashion) during embryogenesis. The generation of cell type diversity during development is based to a great extent on sharp transitions in the dynamical state of embryonic cells, particularly with respect to their developmentally-regulated genes.

2.1. Cell states and dynamics

When a cell divides it inherits not just a set of genes and a particular mixture of molecular components, but it also inherits a dynamical system at a particular *dynamical state*. Dynamical states of many-component systems can be transient, stable, unstable, oscillatory, or chaotic [2]. The cell division cycle in the early stages of frog embryogenesis, for example, is thought to be controlled by a *limit cycle* oscillator [5]. A limit cycle is a continuum of dynamical states that define a stable orbit in the state space surrounding an unstable *node*, a node being a stationary (i.e., time-independent) point, or steady state, of the dynamical system. The fact that cells can inherit dynamical states was demonstrated experimentally by Elowitz and Leibler [6]. These investigators used genetic engineering techniques to provide the bacterium *Escherichia coli* with a set of feedback circuits involving transcriptional repressor proteins such that a biochemical oscillator not previously found in this organism was produced. Individual cells displayed a chemical oscillation with a period longer than the cell cycle. This implied that the dynamical state of the artificial oscillator was inherited across cell generations (if it had not, no periodicity distinct from that of the cell cycle would have been observed). Because the biochemical oscillation was not tied to the cell cycle oscillation, newly divided cells in successive generations found themselves at different phases of the engineered oscillation.

The ability of cells to pass on dynamical states (and not just “informational” macromolecules such as DNA) to their progeny has important implications for developmental regulation, since continuity and stability of a cell’s biochemical identity is key to the performance of its role in

a fully developed organism. Inheritance of alternative cell states that does not depend on sequence differences in inherited genes is called “epigenetic inheritance” [7], and the biological or biochemical states inherited in this fashion are called “epigenetic states.” Although epigenetic states can be determined by reversible chemical modifications of DNA [8], they can also represent alternative steady states of a cell’s network of active or expressed genes [9]. The ability of cells to undergo transitions among a limited number of discrete, stable epigenetic states and to propagate such decisions from one cell generation to the next is essential to the capacity of the embryo to generate diverse cell types.

All dividing cells exhibit oscillatory dynamical behavior in the subspace of the full state space whose coordinates are defined by cell cycle–related molecules. In contrast, cells exhibit alternative stable steady states in the subspace defined by molecules related to states of cell determination and differentiation. During development, a dividing cell might transmit its particular system state to each of its daughter cells, but it is also possible that some internal or external event accompanying cell division could push one or both daughter cells out of the “basin of attraction” in which the precursor cell resided and into an alternative state. (The basin of attraction of a stable node is the region of state space surrounding the node, in which all system trajectories, present in this region, terminate at that node [2].)

It follows that not every molecular species needs to be considered simultaneously in modeling a cell’s transitions between alternative biochemical states. Changes in the concentrations of the small molecules involved in the cell’s housekeeping functions such as energy metabolism and amino acid, nucleotide, and lipid synthesis, occur much more rapidly than changes in the pools of macromolecules such as RNAs and proteins. The latter are indicative of the cell’s gene expression profile and can therefore be considered against an average metabolic background. And even with regard to gene expression profile, most of a cell’s active genes are kept in the “on” state during the cell’s lifetime, since they are also mainly involved in housekeeping functions. The pools of these “constitutively active” gene products can often be considered constant, with their concentrations entering into the dynamic description of the developing embryo as fixed parameters rather than variables. (See Goodwin [10] for an early discussion of separation of timescales in cell activities.)

As mentioned above, it is primarily the regulated genes that are important to consider in analyzing determination and differentiation. And of these

regulated genes, the most important ones for understanding developmental transitions are those whose products control the activity of other genes.

2.2. Epigenetic multistability: the Keller autoregulatory transcription factor network model

Genes are regulated by a set of proteins called *transcription factors*. Like all proteins, these factors are themselves gene products, with the specific function of turning genes on and off. They do this by binding to specific sequences of DNA usually (but not always) “upstream” of the gene’s transcription start site, called the “promoter.” Developmental transitions are frequently controlled by the relative levels of transcription factors [11]. Because the control of most developmentally-regulated genes is a consequence of the synthesis of the factors that regulate their transcription, transitions between cell types during development can be driven by changes in the relative levels of a fairly small number of transcription factors. We can thus gain insight into the dynamical basis of cell type switching (i.e., determination and differentiation) by focusing on molecular circuits, or networks, consisting solely of transcription factors and the genes that specify them. Networks in which the components mutually regulate one another’s expression are termed “autoregulatory.” During development, cells contain a variety of autoregulatory transcription factor circuits. It is obvious that such circuits will have different properties depending on their particular “wiring diagrams” that is the interaction patterns between components.

Transcription factors can be classified as either *activators*, which bind to a site on a gene’s promoter, and enhance the rate of that gene’s transcription over its basal rate, and *repressors*, which decrease the rate of a gene’s transcription when bound to a site on its promoter. The basal rate of transcription depends on constitutive transcription factors, which are distinct from those in the autoregulatory circuits that we consider below. Repression can be competitive or noncompetitive. In the first case, the repressor will interfere with activator binding and can only depress the gene’s transcription rate to the basal level. In the second case, the repressor acts independently of any activator and can therefore potentially depress the transcription rate below basal levels.

Keller [12] used simulation methods to investigate the behavior of several autoregulatory transcription factor networks with a range of wiring diagrams (Fig. 2.1). Each network was represented by a set of n coupled

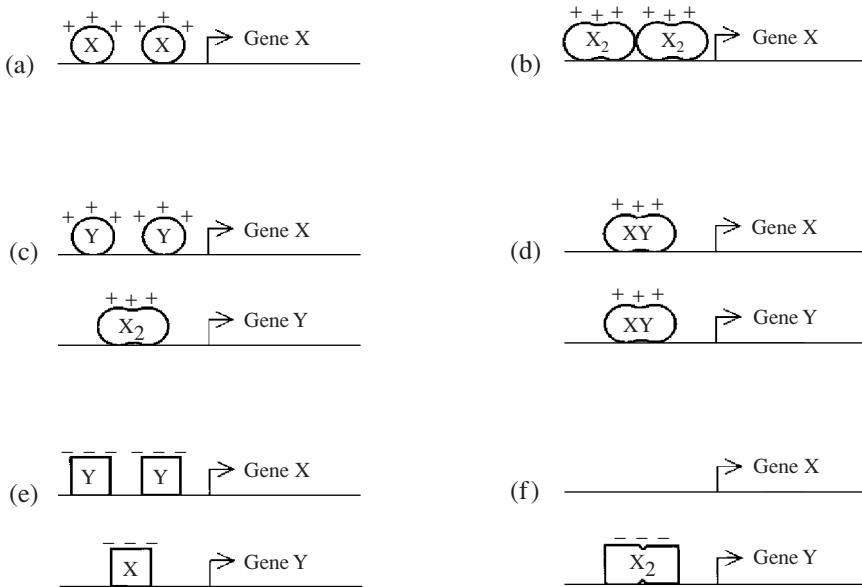


Figure 2.1. Six model genetic circuits discussed by Keller [12]. (a) Autoactivation by monomer X; (b) autoactivation by dimer X_2 ; (c) mutual activation by monomer Y and dimer X_2 ; (d) autoactivation by heterodimer XY; (e) mutual repression by monomers X and Y; (f) mutual repression by dimer X_2 and heterodimer XY. Activation and repression functions are represented respectively by + and -. The various transcription factors bind to the promoters of the genes. Figure used by permission of Elsevier Publishing Co.

ordinary differential equations—one for the concentration of each factor in the network—and the steady-state behaviors of the systems were explored. The questions asked were: how many stationary states exist; are they stable or unstable?

Here we discuss in detail one such network, designated as the “Mutual repression by dimer and heterodimer” (MRDH), shown in Fig. 2.1f. It comprises the gene encoding the transcriptional repressor, X, and the gene encoding the protein, Y, and thus represents a two component network. Below are listed the salient points of the MRDH model.

The rate of synthesis of a transcription factor is proportional to the rate of transcription of the gene encoding that factor. Transcription of a gene, in turn, depends on the specific molecules that are bound to sites in the gene’s promoter. These can be monomeric protein molecules (X, Y), homodimers

(X_2, Y_2) or heterodimers (XY) . Thus a promoter can be in various configurations, with respective relative frequencies, depending on its occupancy. Specifically, in the case of the MRDH network, as characterized by Keller (Fig. 2.1f) the promoter of gene X has no binding site for any activator or repressor molecule; its only configuration is the empty state, whose frequency therefore can be taken as 1. The basal rate of synthesis of X is denoted by S_{X_B} . Protein X is a non-competitive repressor of Y, whose promoter contains a single binding site for X_2 . The monomeric forms of proteins X and Y and the heterodimer XY cannot bind DNA. Thus, while protein Y does not act directly as a transcription factor, it affects transcription since it antagonizes the repressor function of X by interfering with the formation of X_2 . The promoter of gene Y can therefore be in two configurations: its binding site for X_2 is either occupied or not, with respective relative frequencies $K_X K_{X_2} [X]^2 / (1 + K_X K_{X_2} [X]^2)$ and $1 / (1 + K_X K_{X_2} [X]^2)$. Here K_X and K_{X_2} are, respectively, the binding affinity of X_2 to the promoter of gene Y and the dimerization rate constant for the formation of X_2 (we used the relationship $K_X [X_2] = K_X K_{X_2} [X]^2$). Synthesis of Y in both configurations is activator-independent with a rate denoted by S_{Y_B} . To incorporate the fact that X_2 reduces the rate of transcription of Y in a non-competitive manner, in the occupied Y promoter configuration S_{Y_B} is replaced with ρS_{Y_B} , with $\rho \leq 1$.

The overall transcription rate of a gene is calculated as the sum of products. Each term in the sum corresponds to a particular promoter occupancy configuration and is represented as a product of the frequency of that configuration and the rate of synthesis resulting from that configuration. In the MRDH network this rate for gene X is thus $1 \times S_{X_B}$, because it has only a single (empty) promoter configuration. The promoter of gene Y can be in two configurations (with rates of synthesis S_{Y_B} and ρS_{Y_B} , see above), therefore its overall transcription rate is $(1 + \rho K_X K_{X_2} [X]^2) S_{Y_B} / (1 + K_X K_{X_2} [X]^2)$.

The rate of decay of a transcription factor is a sum of terms, with each being proportional to the concentration of a particular complex in which the transcription factor participates. This is equivalent to assuming exponential decay. For the transcriptional repressor X in the MRDH network these complexes include the monomer X, the homodimer X_2 and the heterodimer XY. Denoting the corresponding decay constants as d_X , d_{X_2} and d_{XY} , the overall decay rate of X is given by $d_X [X] + 2d_{X_2} K_{X_2} [X]^2 + d_{XY} K_{XY} [X][Y]$, with K_{XY} being the rate constant for the formation of the heterodimer

XY (i.e., $[XY] = K_{XY}[X][Y]$). The analogous quantity for protein Y is $d_Y[Y] + d_{XY}K_{XY}[X][Y]$.

With the above ingredients, the steady-state concentrations of X and Y in the MRDH network, are determined by

$$\frac{d[X]}{dt} = S_{X_B} - \{d_X[X] + 2d_{X_2}K_{X_2}[X]^2 + d_{XY}K_{XY}[X][Y]\} = 0 \quad (2.1)$$

$$\frac{d[Y]}{dt} = \frac{1 + \rho K_X K_{X_2} [X]^2}{1 + K_X K_{X_2} [X]^2} S_{Y_B} - \{d_Y[Y] + d_{XY}K_{XY}[X][Y]\} = 0. \quad (2.2)$$

Keller found that if in the absence of the repressor X the rate of synthesis of protein Y is high, in its presence, the system described by Eqs. (2.1) and (2.2) exhibits three steady states, as shown in Fig. 2.2. Steady states 1 and 3 are stable, thus could be considered as defining two distinct cell types, while

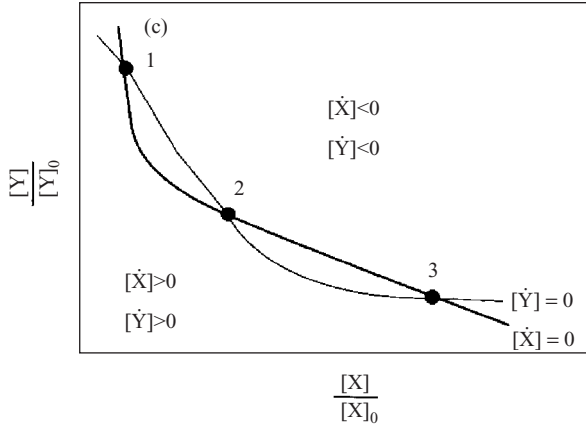


Figure 2.2. The solutions of the steady-state Eqs. 1 and 2, given in terms of the steady state solutions $[X_0]$ and $[Y_0]$. Here $[X_0]$ is defined as the steady state cellular level of monomer X produced in the presence of steady state cellular level $[Y_0]$ of monomer Y by the rate of transcription $[S_{X_0}]$. Thus by definition (see Eq. 1), $S_{X_0} = d_X[X_0] + 2 d_{X_2} K_{X_2} [X_0]^2 + d_{XY} K_{XY} [X_0] [Y_0]$. Since along the thick and thin lines, respectively, $d[X]/dt \equiv [X^Y] = 0$ and $d[Y]/dt \equiv [Y^Y] = 0$, the intersections of these curves correspond to the steady state solutions of the system of equations, Eqs. 1 and 2. Steady states 1 and 3 are stable, while steady state 2 is unstable. From Keller (1995) [12]. Used by permission of Elsevier Publishing Co.

steady state 2 is unstable. In an example using realistic kinetic constants, the steady-state values of $[X]$ and $[Y]$ at the two stable steady states differ substantially from one another, showing that the dynamical properties of these autoregulatory networks of transcription factors can provide the basis for generating stable alternative cell fates during early development.

The biological validity of Keller's model depends on whether switching between these alternative states is predicted to occur under realistic conditions. By this criterion the model works: the microenvironment of a cell, containing an autoregulatory network of transcription factors, could readily induce changes in the rate of synthesis of one or more of the factors via signal transduction pathways that originate outside the cell ("outside-in" signaling [13]). Moreover, the microenvironment can also affect the activity of transcription factors in an autoregulatory network by indirectly interfering with their localization in the cell's nucleus, where transcription takes place [14]. In addition, cell division may perturb the cellular levels of autoregulatory transcription factors, particularly if they or their mRNAs are unequally partitioned between the daughter cells. Any jump in the concentration of one or more factors in the autoregulatory system can bring it into a new basin of attraction and thereby lead to a new stable cell state.

2.3. Dependence of differentiation on cell-cell interaction: the Kaneko-Yomo "isologous diversification" model

The Keller model shows that the existence of multiple steady states in an embryonic cell's state space makes it possible, in principle, for more than one cell type to arise among its descendants. However this capability does not, by itself, provide the conditions under which such a potentially divergent cell population would actually be produced and persist as long as it is needed.

Experimental observations suggest that cell differentiation depends on properties of multicellular aggregates rather than simply those of individual cells. For example, during muscle differentiation in the early frog embryo, the muscle precursor cells must be in contact with one another throughout gastrulation (the set of rearrangements that establish the body's main tissue layers) in order to develop into terminally differentiated muscle [15, 16].

The need for cells to act in groups in order to acquire new identities during development has been termed the "community effect" [15]. This

phenomenon is a developmental manifestation of the general property of cells and other dynamical systems of assuming one or another of their possible internal states in a fashion that is dependent on inputs from their external environment. In the case noted above the external environment consists of other cells of the same genotype.

Kaneko, Yomo and co-workers [17-20] have described a previously unknown chemical-dynamic process, termed “isologous diversification,” by which replicate copies of the same dynamical system (e.g., cells of the same initial type) can undergo stable differentiation simply by virtue of exchanging chemical substances with one another. This differs from the model described above in that the final state achieved exists only in the phase space of the collective “multicellular” system. Whereas the distinct local states of each cell within the collectivity are mutually reinforcing, these local states are not necessarily attractors of the dynamical system representing the individual cell, as they are in Keller’s model. The Kaneko-Yomo system thus provides a model for the community effect.

The following is a simple version of the model, based on Kaneko and Yomo [17]. Improvements and generalizations of the model presented in subsequent publications [18-20] do not change its qualitative features.

Kaneko and Yomo consider a system of originally identical cells with intra- and inter-cell dynamics, which incorporate cell growth, cell division and cell death. The dynamical variables are the concentrations of molecular species (“chemicals”) inside and outside the cells. The criterion by which differentiated cells are distinguished is the average of the intracellular concentrations of these chemicals (over the cell cycle). As a vast simplification only three chemicals, A , B and S , with respective time-dependent intracellular ($x_i^A(t)$, $x_i^B(t)$, $x_i^S(t)$ in the i th cell) and intercellular ($X^A(t)$, $X^B(t)$, $X^S(t)$) concentrations are considered in each cell and the surrounding medium. One of those (S) serves as source for the others. The model has the following features:

The source chemical S is catalyzed by a constitutive enzyme to produce chemical A , which in turn is catalyzed by a regulated enzyme to produce the chemical B . Chemical B on one hand is catalyzed by its own regulated enzyme to produce A , on the other hand controls the synthesis of DNA. This sequence of events is schematically shown in Fig. 2.3. The concentration of the constitutive enzyme is assumed to have the same constant value E^S in each cell, whereas those of the regulated enzymes in the i th cell, E_i^A and E_i^B are both taken to be proportional to the concentration x_i^B of the

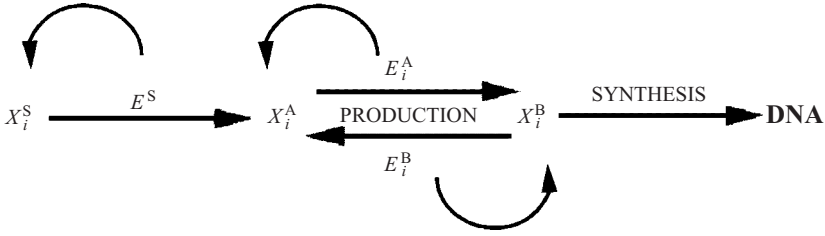


Figure 2.3. Schematic representation of the intracellular dynamics used in the model of Kaneko and Yomo [17]. Curved arrows symbolize catalysis. The variables $x_i^A(t)$, $x_i^B(t)$, $x_i^S(t)$ and E_i^A , E_i^B , E^S denote respectively the concentrations of chemicals A, B and S and their enzymes in the i -th cell, as explained in the text.

chemical B (and therefore dependent on time) in that cell, $E_i^A = e_A x_i^B$ and $E_i^B = e_B x_i^A$ (e_A and e_B are constants). Thus, in terms of chemicals A and B the intracellular dynamics is described by

$$\begin{aligned} \frac{dx_i^A}{dt} &= (e_B x_i^B) x_i^B - (e_A x_i^B) x_i^A + E^S x_i^S \\ \frac{dx_i^B}{dt} &= (e_A x_i^A) x_i^A - (e_B x_i^B) x_i^B - k x_i^B \end{aligned} \quad (2.3)$$

Here the factor k accounts for the decrease of B due to its role in DNA synthesis (see Fig. 2.3). Note the non-linear character of these equations.

Cells are assumed to interact with each other through their effect on the intercellular concentrations of the chemicals A and B. Chemicals are transported in and out of cells. The rate of their transport into a cell is proportional to their concentration outside, but also depends on the internal state of the cell, as characterized by the intracellular concentrations of A and B. Kaneko and Yomo assume that the rate of import of chemical M (i.e., A, B or S) into the i th cell, denoted by $Transp_i^M$, has the form

$$Transp_i^M(t) = p (x_i^A + x_i^B)^3 X^M \quad (2.4)$$

Here p is a constant. As long as the dependence of $Transp$ on the intracellular concentrations is nonlinear, the choice of the exponent (taken to be 3 above) leads to the same qualitative result.

Besides active transport described by Eq. (2.4) chemicals also enter the cells by diffusion through the membrane. The corresponding rate is

taken as

$$Diff_i^M(t) = D [X^M(t) - x_i^M(t)] \quad (2.5)$$

where D is a (diffusion) constant.

Combining intracellular (Eq. 2.3) and intercellular (Eqs. 2.4 and 2.5) dynamics, the rate equations for the intracellular chemicals become

$$\begin{aligned} \frac{dx_i^S}{dt} &= -Ex_i^S + Transp_i^S + Diff_i^S \\ \frac{dx_i^A}{dt} &= (e_B x_i^B) x_i^B - (e_A x_i^B) x_i^A + Ex_i^S + Transp_i^A + Diff_i^A \\ \frac{dx_i^B}{dt} &= (e_A x_i^B) x_i^A - (e_B x_i^B) x_i^B - kx_i^B + Transp_i^B + Diff_i^B \end{aligned} \quad (2.6)$$

It is further assumed that only the source chemical is supplied by a flow from an external tank to the chamber containing the cells. Since it must be transported across the cell membrane to produce chemical A (Eqs. 2.4 and 2.5), the intercellular dynamics of the source chemical is described by

$$\frac{dX^S}{dt} = f (\overline{X^S} - X^S) - \sum_{i=1}^N (Transp_i^S + Diff_i^S) \quad (2.7)$$

Here $\overline{X^S}$ is the concentration of the source chemical in the external tank, f is its flow rate into the chamber and N is the total number of cells in the system.

Kaneko and Yomo [17] consider cell division to be equivalent to the accumulation of a threshold quantity of DNA. DNA is synthesized from chemical B and therefore the i th cell, born at time t_i^0 , will divide at $t_i^0 + T$ (T defines the cell cycle time) when the amount of B in its interior (proportional to x_i^B) reaches a threshold value. Mathematically this condition is expressed as $\int_{t_i^0}^{t_i^0+T} dt x_i^B(t) \geq R$ in the model, with R being the threshold value.)

To avoid infinite growth in cell number, a condition for cell death has to also be imposed. It is assumed that a cell will die if the amount of chemicals A and B in its interior is below the “starvation” threshold S (which is expressed as $\lfloor x_i^A(t) + x_i^B(t) \rfloor < S$).

Simulations based on the above model and its generalizations, using a larger number of chemicals [18-20], led to the following general features, which are likely to pertain to real, interacting cells as well:

1. As the model cells replicate (by division) and interact with one another, eventually multiple biochemical states corresponding to distinct cell types appear. The different types are related to each other by a hierarchical structure in which one cell type stands at the apex, cell types derived from it stand at subnodes, and so on. Such pathways of generation of cell type, which are seen in real embryonic systems, are referred to as developmental lineages.
2. The hierarchical structure appears gradually. Up to a certain number of cells (which depends on the model parameters), all cells have the same biochemical state (i.e., $x_i^A(t)$, $x_i^B(t)$ and $x_i^S(t)$ are independent of i). When the total number of cells rises above a certain threshold value, the state with identical cells is no longer stable. Small differences between cells first introduced by random fluctuations in chemical concentrations start to be amplified. For example, synchrony of biochemical oscillations in different cells of the cluster may break down (by the phases of $x_i^A(t)$, $x_i^B(t)$, $x_i^S(t)$ becoming dependent on i) (Fig. 2.4a). Ultimately, the population splits into a few groups (“dynamical clusters”), with the phase of the oscillator in each group being offset from that in other groups, like groups of identical clocks in different time zones.
3. When the ratio of the number of cells in the distinct clusters falls within some range (depending on model parameters), the differences in intracellular biochemical dynamics are mutually stabilized by cell-cell interactions.
4. With further increase of cell number, the average concentrations of the chemicals over the cell cycle become different (Fig. 2.4b). That is to say, groups of cells come to differ not only in the phases of the same biochemical oscillations, but also in their average chemical composition integrated over the entire lifetimes of the cells. After the formation of cell types, the chemical compositions of each group are inherited by their daughter cells (Fig. 2.4c).

In contrast to the Keller model described above, in which different cell types represent a choice among basins of attraction for a multi-attractor system, with external influences having the potential to bias such preset alternatives, in the Kaneko-Yomo model interactions between cells can give rise to stable intracellular states which would not exist without such interactions. Isologous diversification thus provides a plausible model for

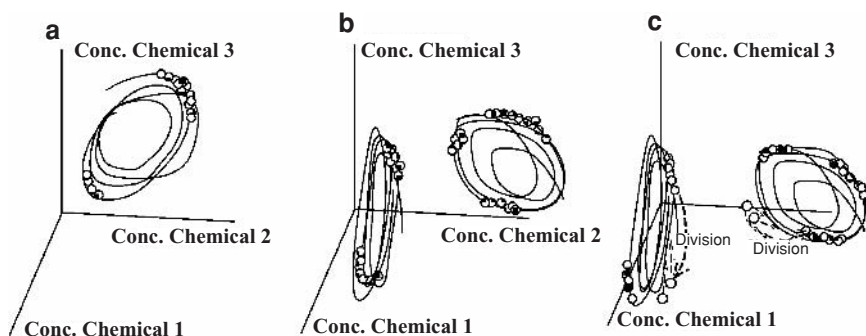


Figure 2.4. Schematic representation of some stages of the differentiation scenario in the model of Kaneko and Yomo [17]. Stage 1 (not shown): Synchronous divisions with synchronous oscillations of the chemicals. Up to a certain number of cells dividing cells arising from a single precursor cell have the same characteristics. Although each cell division is not exactly identical due to the accompanying fluctuation in the biochemical composition, the phase of oscillation in the concentrations, and as a result, the timing of cell division, remains synchronous for all cells; (a) Stage 2: Clustering in the phases of oscillations. When the number of cells rises above a certain (threshold) value, the state with identical cells is no longer stable. Small differences introduced by fluctuations start to be amplified, until the synchrony of the oscillations is broken. The cells then split into a few groups, each having a different oscillation phase. The cells within each group are identical in phase. This diversification in the phases, however, is not equivalent to cell differentiation, because the time average of the biochemical concentrations reveals that the cells are almost identical; (b) Differentiation in chemical composition. With the further increase of the cell number, the average concentrations of the biochemicals over the cell cycle become different. The orbits of chemical dynamics plotted in the phase space of biochemical concentrations lie in distinct regions within the phase space, while the phases of oscillations for cells within each group remain different; (c) “Breeding true” of the differentiated cells. This is indicated by new system points, generated by cell division, residing on trajectories (dashed lines) that track the parental trajectories. After the formation of cell types, the chemical compositions of each group are inherited by their daughter cells. In other words, chemical compositions of cells are recursive over subsequent divisions. Adapted, with changes, from Kaneko (2003) [20], with permission.

the community effect [15], described above. It is reasonable to expect that both intrinsic multistability of a dynamical system of the sort analyzed by Keller, and interaction-dependent multistability, as described by Kaneko, Yomo, and coworkers, based as they are on generic properties of complex dynamical systems, are utilized in initiating developmental decisions in various contexts in different organisms.

3. Biochemical Oscillations and Segmentation

A wide variety of animal types, ranging across groups as diverse as insects, annelids (e.g., earthworms), and vertebrates, undergo *segmentation* early in development, whereby the embryo, or a major portion of it, becomes subdivided into a series of tissue modules [1]. These modules typically appear similar to each other when initially formed; later they may follow distinct developmental fates and the original segmental organization may be all but obscured in the adult form. Somite formation (or somitogenesis) is a segmentation process in vertebrate embryos in which the tissue to either side of the central axis of the embryo (where the backbone will eventually form) becomes organized into parallel blocks of tissue.

Somitogenesis takes place in a sequential fashion. The first somite begins forming as a distinct cluster of cells in the anterior region (towards the head) of the embryo's body. Each new somite forms just posterior (towards the tail) to the previous one, budding off from the anterior portion of the unsegmented presomitic mesoderm (PSM) (Fig. 2.5). Eventually, 50 (chick), 65 (mouse), or as many as 500 (certain snakes) of these segments will form.

3.1. Oscillatory dynamics and somitogenesis

In the late 19th century the biologist William Bateson speculated that the formation of repetitive blocks of tissue, such as the somites of vertebrates or the segments of earthworms might be produced by an oscillatory process inherent to developing tissues [21]. More recently, Pourquié and coworkers made the significant observation that the gene *c-hairy1*, which specifies a transcription factor (see previous section), is expressed in the PSM of avian embryos in cyclic waves whose temporal periodicity corresponds to the formation time of one somite [22, 23] (Fig. 2.5). The *c-hairy1* mRNA and protein product are expressed in a temporally-periodic fashion in individual cells, but since the phase of the oscillator is different at different points along the embryo's axis, the areas of maximal expression sweep along the axis in a periodic fashion.

Experimental evidence suggests that somite boundaries form when cells which have left a posterior growth zone move sufficiently far away from a source of a diffusible protein known as fibroblast growth factor 8 (FGF8) in the tailbud at the posterior end of the embryo [24]. The FGF gradient thus

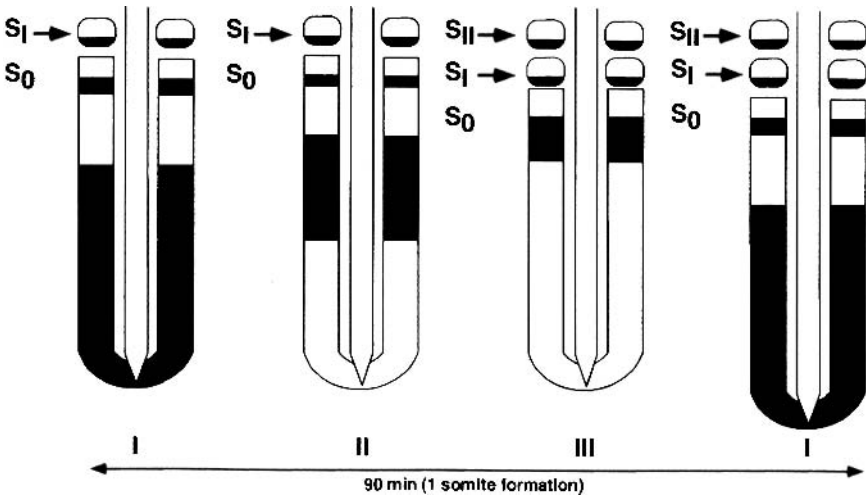


Figure 2.5. Formation of somites, segmented blocks of tissue along the main body axis, in chicken embryos, associated with traveling waves of expression of a regulatory protein (*c-hairy1*). The protein distribution is represented as black regions in this schematic drawing. *c-hairy1* is expressed in a temporally-periodic fashion in individual cells, but since the phase of the oscillator is different at different points along the embryo's axis the areas of maximal expression sweep along the axis in a periodic fashion. Expression is confined to the caudal (toward the tail) half of each somite, where it plays a functional role in causing separation from adjacent, presegmented tissue. S_0 , S_1 and S_{II} represent successively-forming somites. Reprinted, with modifications, from *Cell*, Vol. 91, I. Palmeirim, D. Henrique, D. Ish-Horowicz & O. Pourquié, Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis [22], p. 642, Copyright 1997, with permission from Elsevier.

acts as a “gate” that, when its low end coincides with a particular phase of the segmentation clock, results in formation of a boundary [24, 25]. The general features of this mechanism (called the “clock and wavefront” model) were predicted on the basis of dynamical principles two decades before there was any direct evidence for a somitic oscillator [26].

3.2. The Lewis model of the somitogenesis oscillator

During somitogenesis in the zebrafish a pair of transcription factors known as *her1* and *her7*, which are related to chicken *hairy1* (see above),

oscillate in the PSM in a similar fashion, as does the cell surface signaling ligand deltaC. Lewis [27] and Monk [28] have separately suggested that *her1* and *her7* constitute an autoregulatory transcription factor gene circuit of the sort treated by Keller [12] (see section 2, above), and are the core components of the somitic oscillator in zebrafish. Lewis also hypothesized that deltaC, whose signaling function is realized by activating the Notch receptors on adjacent cells, is a downstream effector of this oscillation. The two *her* genes negatively regulate their own expression [29, 30] and are positively regulated by signaling via Notch [20, 31]. Certain additional experimental results [32] led Lewis to the conclusion that signaling by the Notch pathway, usually considered to act in the determination of cell fate [33], in this case acts to keep cells in the segment-generating growth zone in synchrony [27]. Such synchrony has been experimentally confirmed in chicken embryos [34, 35].

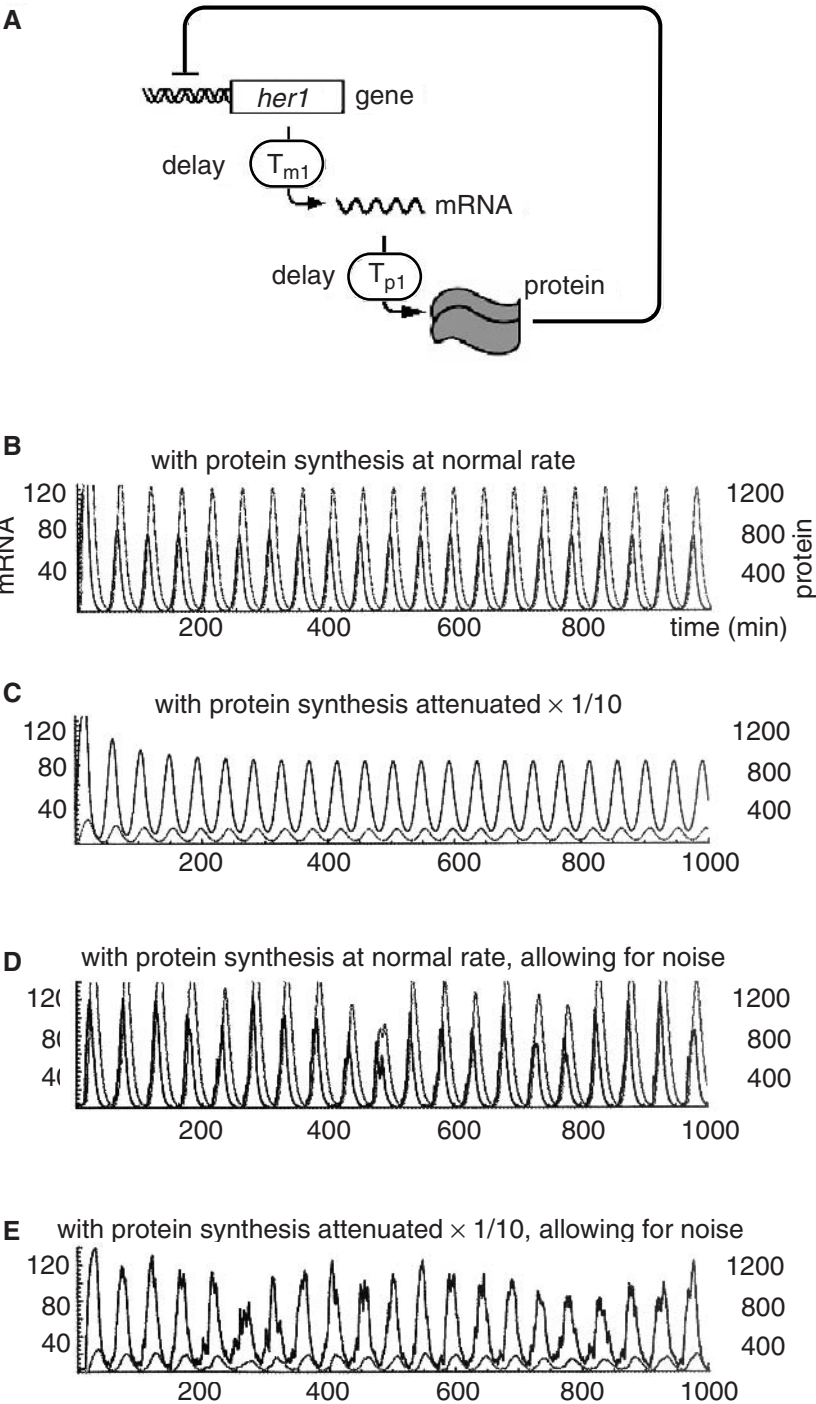
Lewis provides a simple mechanism for the oscillatory expression of the *her1* and *her7* genes, which we briefly summarize here. The model is based on the reasonable assumption that there exists a feedback loop in which the Her1 and Her7 proteins directly bind to the regulatory DNA of their own genes to inhibit transcription. Also incorporated into the model is the recognition that there is always a delay between the initiation of transcription and the initiation of translation, T_m (since it takes time for the mRNA molecule to translocate into the cytoplasm), as well as between the initiation of translation and the emergence of a complete functional protein molecule, T_p .

For a given autoregulatory gene, let $m(t)$ be the number of mRNA molecules in a cell at time t and let $p(t)$ be the number of the corresponding protein molecule. The rate of change of m and p , are then assumed to obey the following equations

$$\frac{dp(t)}{dt} = am(t - T_p) - bp(t) \quad (3.1)$$

$$\frac{dm(t)}{dt} = f[p(t - T_p)] - cm(t). \quad (3.2)$$

Here the constants b and c are the decay rates of the protein and its mRNA, respectively, a is the rate of production of new protein molecules and $f(p)$ is the rate of production of new mRNA molecules. The function $f(p)$ is assumed to be a decreasing function of the amount of protein. (The form Lewis and Monk used is $f(p) = k/(1 + p^2/p_0^2)$, with constants



k and p_0 , to represent the action of an inhibitory protein, assumed to be a dimer. However, results were surprisingly insensitive to the specific form of $f(p)$.)

The above delay differential equations were numerically solved for *her1* and *her7* (for which Lewis was able to estimate the values of all the model parameters in Eqs. 3.1 and 3.2 using experimental results). The solutions indeed exhibit sustained oscillations in the concentration of Her1 and Her7, with the predicted periods close to the observed ones (Fig. 2.6). The important conclusions from the analysis of Lewis are that no oscillations are possible if delay is not incorporated (i.e., $T_m = T_p = 0$) and that the oscillators are quite insensitive to blockade of protein synthesis (i.e., to the value of a in Eq. 3.1). Furthermore, Lewis showed that incorporation into the model of the inherently noisy nature of gene expression (by adding stochastic effects to the deterministic equations, Eqs. 3.1 and 3.2) reinforces continued oscillations. (Without noise oscillations are eventually damped, which would upset normal somite formation beyond the first few.)

Figure 2.6. Cell autonomous gene expression oscillator. A. Molecular control circuitry for a single gene, *her1*, whose protein product acts as a homodimer to inhibit *her1* expression. In case of a pair of genes (i.e. *her1* and *her7*) the analogous circuit would contain an additional branch with coupling between the two branches. B. Computed behavior for the system in A (defined by Eqs. 8 and 9) in terms of the number mRNA molecules per cell in black and protein molecules in gray. Parameter values were chosen appropriate for the *her1* homodimer oscillator (see the form of the function $f(p)$) based on experimental results: $a = 4.5$ protein molecules per mRNA molecules per minute; $b = c = 0.23$ molecules per minute, corresponding to protein and mRNA half-lives of 3 minutes; $k = 33$ mRNA per diploid cell per minute, corresponding to 1000 transcripts per hour per gene copy in the absence of inhibition; $\mu_0 = 40$ molecules, corresponding to a critical concentration of around 10^{-9} M in a $5 \mu\text{m}$ diameter cell nucleus; $T_m \approx 20.8$ min; $T_p \approx 2.8$ min. C. Decreasing the rate of protein synthesis ($a = 0.45$) causes little or no effect in the period of oscillation. All the other parameters are the same as in A. D. Computed behavior for system in A when the noisy nature of the gene expression is taken into account. To model stochastic effects Lewis introduced one more independent parameter, the rate constant k_{off} for the dissociation of the repressor protein (i.e., Her1) from its binding site on the regulatory DNA of its own gene (*her1*). Results are shown for $k_{off} = 1 \text{ min}^{-1}$, corresponding to a mean lifetime of 1 min of the repressor bound state. E. The same as in D except for the rate of protein synthesis, which is as in C. The parameter values not mentioned explicitly in C-E are the same as in B. Protein concentration is represented by the upper curve B and D and the lower curve in C and E. Adapted, with changes, from Lewis (2003) [27], with permission.

4. Reaction-Diffusion Mechanisms and Embryonic Pattern Formation

The nonuniform distribution of any chemical substance, whatever the mechanism of its formation, can clearly provide spatial information to cells. For at least a century embryologists have considered models for pattern formation and its regulation that employ diffusion gradients [1]. Only in the last decade, however, has convincing evidence been produced that this mechanism is utilized in early development. The prime evidence comes from studies of mesoderm *induction*, a key event preceding gastrulation in the frog *Xenopus*. Nieuwkoop [36] originally showed that mesoderm (the middle of the three “germ layers” in the three-layered gastrula—the one that gives rise to muscle, skeletal tissue, and blood) only appeared when tissue from the upper half of an early embryo (“animal cap”) was juxtaposed with tissue from the lower half of the embryo (“vegetal pole”). By themselves, animal cap and vegetal pole cells, respectively, only produce ectoderm, which gives rise to skin and nervous tissue, and endoderm, which gives rise to the intestinal lining. Later it was found that several released, soluble factors of the TGF- β protein superfamily and the FGF protein family could substitute for the inducing vegetal pole cells (reviewed by Green [37]). Both TGF- β [38] and FGFs [39] can diffuse over several cell diameters.

None of this proves beyond question that simple diffusion of such released signal molecules (called “morphogens”) between and among cells, rather than some other, cell-dependent mechanism, actually establishes the gradients in question. Kerszberg and Wolpert [40] for example, assert that capture of morphogens by receptors impedes diffusion to an extent that stable gradients can never arise by this mechanism. They propose that morphogens are instead transported across tissues by a “bucket brigade” mechanism in which a receptor-bound morphogen on one cell moves by being handed off to receptors on an adjacent cell.

Lander and co-workers [41] use quantitative estimates of the spreading of morphogens in an insect developmental system [42, 43] and conclude, using a mathematical model of morphogen spread and reversible binding to receptors, that free diffusion is indeed a plausible physical mechanism for establishing embryonic gradients. The model of Lander et al. [41], as well as more complex ones describing formation of the embryo’s primary (anteroposterior) axis and generation of left-right asymmetry can be

considered examples of a generalized reaction-diffusion system, which we will now describe.

4.1. Reaction-diffusion systems

The rate of change in the concentrations of n interacting molecular species (c_i , $i = 1, 2, \dots, n$) is determined by their reaction kinetics and expressed in terms of ordinary differential equations

$$\frac{dc_i}{dt} = F_i(c_1, c_2 \dots c_n). \quad (4.1)$$

The explicit form of the functions F_i in Eq. (4.1) depends on the details of the reactions. Spatial inhomogeneities also cause time variations in the concentrations even in the absence of chemical reactions. If these inhomogeneities are governed by diffusion, then in one spatial dimension,

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial x^2}. \quad (4.2)$$

Here D_i is the diffusion coefficient of the i th species. In general, both diffusion and reactions contribute to the change in concentration and the time dependence of the c_i s is governed by reaction-diffusion equations

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial x^2} + F_i(c_1, c_2 \dots c_n). \quad (4.3)$$

Reaction-diffusion systems exhibit characteristic parameter-dependent bifurcations (the “Turing instability”), which are thought to serve as the basis for pattern formation in several embryonic systems, including butterfly wing spots [44], stripes on fish skin [45], distribution of feathers on the skin of birds [46], the skeleton of the vertebrate limb [47, 48], and the primary axis of the developing vertebrate embryo [49], which we discuss in detail, below.

4.2. Axis formation and left-right asymmetry

Gastrulation in the frog embryo is initiated by the formation of an indentation, the “blastopore,” through which the surrounding cells invaginate, or tuck into the hollow blastula. Spemann and Mangold [50] discovered that the anterior blastopore lip constitutes an *organizer*: a population of cells that directs the movement of other cells. The action of the

Spemann-Mangold organizer ultimately leads to the formation of the notochord, the rod of connective tissue that first defines the anteroposterior body axis, and to either side of which the somites later form (see Section 3, above). These investigators also found that an embryo with an organizer from another embryo at the same stage transplanted at some distance from its own organizer would form two axes, and conjoined twins would result. Other classes of vertebrates have similarly acting organizers.

A property of this tissue is that if it is removed, adjacent cells differentiate into organizer cells and take up its role. This indicates that one of the functions of the organizer is to suppress nearby cells with similar potential from exercising it. This makes the body axis a partly self-organizing system. The formation of the body axis in vertebrates also exhibits another unusual feature: while it takes place in an apparently symmetrical fashion, with the left and right sides of the embryo seemingly equivalent to one another, at some point the symmetry is broken. Genes such as *nodal* and *lefty* start being expressed differently on the two sides of the embryo [51], and the whole body eventually assumes a partly asymmetric morphology, particularly with respect to internal organs, such as the heart.

Turing [52] first demonstrated that reaction-diffusion systems like that represented in Eq. (4.3) will, with appropriate choice of parameters and boundary conditions, generate self-organizing patterns, with a particular propensity to exhibit symmetry breaking across more than one axis. Using this class of models, Meinhardt [49] has presented an analysis of axis formation in vertebrates and the breaking of symmetry around these axes.

4.3. Meinhardt's models for axis formation and symmetry breaking

The first goal a model of axis formation has to accomplish is to generate an organizer de novo. For this high local concentrations and graded distributions of signaling molecules are needed. This can be accomplished by the coupling of a self-enhancing feedback loop acting over a short range with a competing inhibitory reaction acting over a longer range. The simplest system that can produce such a molecular pattern in the x-y plane consists of a positively autoregulatory activator (with concentration $A(x, y; t)$) and an inhibitor (with concentration $I(x, y; t)$). The activator controls the production of the inhibitor, which in turn limits the production of the activator.

This process can be described by the following reaction-diffusion system [49]

$$\frac{\partial A}{\partial t} = D_A \left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} \right) + s \frac{A^2 + I_A}{I (1 + s_A A^2)} - k_A A \quad (4.4a)$$

$$\frac{\partial I}{\partial t} = D_I \left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \right) + s A^2 - k_I I + I_I \quad (4.4b)$$

The A^2 terms specify that the feedback of the activator on its own production and that of the inhibitor in both cases is non-linear. The factor $s > 0$ describes positive autoregulation, the capability of a factor to induce positive feedback on its own synthesis. This may occur by purely chemical means (“autocatalysis”), which is the mechanism assumed by Turing [52] when he first considered systems of this type. More generally, in living tissues, positive autoregulation occurs if a cell’s exposure to a factor it has secreted causes it to make more of the same factor [53]. The inhibitor slows down the production of the activator (i.e., the $1/I$ factor in the second term in Eq. 4.4a). Both activator and inhibitor diffuse (i.e., spread) and decay with respective diffusion (D_A , D_I) and rate constants (k_A , k_I). The small baseline inhibitor concentrations, I_A and I_I can initiate activator self-enhancement or suppress its onset, respectively, at low values of A . The factor s_A , when present, leads to saturation of positive autoregulation. Once the positive autoregulatory reaction is under way, it leads to a stable, self-regulating pattern in which the activator is in dynamic equilibrium with the surrounding cloud of the inhibitor.

The various organizers and subsequent inductions leading to symmetry breaking, axis formation and the appearance of the three germ layers in amphibians during gastrulation, can all, in principle, be modeled by the reaction-diffusion system in Eqs. (4.4a) and (4.4b), or by the coupling of several such systems. The biological relevance of such reaction-diffusion models depends on whether there exist molecules that can be identified as activator-inhibitor pairs. Meinhardt’s model starts with a default state, which consists of ectoderm. Patch-like activation generates the first “hot spot”, the *vegetal pole organizer*, which induces endoderm formation (simulation in panel A in Fig. 2.7). A candidate for the diffusible activator in the corresponding self-enhancing loop for endoderm specification is the TGF- β -like factor *Derriere*, which activates the VegT transcription factor [54].

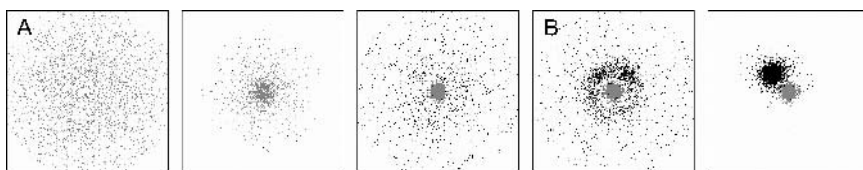


Figure 2.7. Axial pattern formation and induction of two “hot spots” in Meinhardt’s model of axis formation. (A) The interaction of a short ranging positive feedback loop (activator, gray) and a long ranging inhibitory substance (not shown) constitutes an unstable system. In the simulation shown a small initial elevation of the activator leads to a focal activation. In a system without pre-localized determinants, such a reaction could be responsible for the formation of the vegetal pole. (B) A second such system (black) forms a second hot spot next to the first if it is activated over a long range and locally repressed by the first activator. This process can lead to symmetry breaking. According to the model, this corresponds to the Nieuwkoop center at a position displaced from the pole. Adapted, with changes from Meinhardt (2001) [49]. Used with permission.

VegT expression remains localized to the vegetal pole, but not because of lack of competence of the surrounding cells to produce VegT [55]. These findings provide circumstantial evidence for the existence of the inhibitor required by the reaction-diffusion model. Subsequently, a second feedback loop forms a second hot spot in the vicinity of the first, in the endoderm. This is identified with the “Nieuwkoop center,” a second organizing region, which appears in a specific quadrant of the blastula (see Fig. 2.7). A candidate for the second self-enhancing loop is FGF together with Brachyury [56]. Interestingly, the inhibitor for this loop is hypothesized to be the first loop itself (i.e., the vegetal pole organizer), which acts as local repressor for the second. As a result of this local inhibitory effect, the Nieuwkoop center is displaced from the pole (simulation in panel B in Fig. 2.7). With the formation of the Nieuwkoop center the spherical symmetry of the embryo is broken. In Meinhardt’s model this symmetry breaking “propagates” and thus forms the basis of further symmetry breakings, in particular the left-right asymmetry (see below).

By secreting several diffusible factors, the Nieuwkoop center induces the formation of the Spemann-Mangold organizer [57]. (If the second feedback loop, responsible for the Nieuwkoop center is not included in the model, two Spemann-Mangold organizers appear symmetrically with respect to the animal-vegetal axis and no symmetry breaking occurs.) With the formation of the Spemann-Mangold organizer the developmental process is in

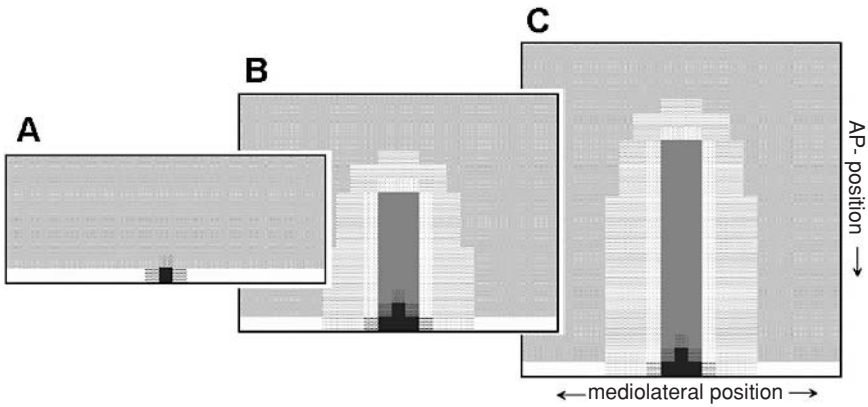


Figure 2.8. Formation of the midline and enfolding of the anteroposterior axis in Meinhardt's model of axis formation. In this simplified simulation a system that is tuned to make stripes (dark gray) is triggered by the organizer, i.e., a system that is activated in a spot-like manner (black). Because the stripe system (corresponding to the notochord) also repels the spot system (corresponding to the Spemann organizer), the spot system is shifted in front of the tip of the stripe, causing its straight elongation. Cells therefore temporarily acquire organizer quality before participating in midline formation. Due to saturation in the self-enhancement, the stripe system does not disintegrate into individual patches, but instead establishes the midline. This, in turn, generates positional information for the dorsoventral or mediolateral axis by acting as a sink for a ubiquitously produced substance (medium and light gray, e.g., BMP-4). The local concentration of the latter is a measure of the distance from the midline. See Meinhardt (2001) [49] for additional details. (Figure adapted with changes from Meinhardt (2001) [49]. Used with permission.)

full swing. The organizer triggers gastrulation, in the course of which the germ layers acquire their relative positions and the notochord forms. This long thin structure marks the midline of the embryo, which itself inherits organizer function and eventually establishes the primary (AP) embryonic axis. A simulation of midline formation, based on Meinhardt's model is shown in Fig. 2.8.

Finally, the breaking of left-right symmetry can be understood again as a competition between already existing and newly developing self-enhancing loops, similarly to the formation of the Nieuwkoop center. The molecule that best fulfils the role of the "left" activator in Meinhardt's model is the product of the *nodal* gene, which is a diffusible, positively autoregulatory member of the TGF- β superfamily. Nodal induces expression from the embryonic midline of another TGF- β related molecule, Lefty, and Lefty

antagonizes Nodal production [58-60]. Because Nodal and Lefty are antagonistic diffusible signals that differ in the range of their activities [59, 61, 62], the ingredients for a symmetry breaking event along the primary embryonic axis are present [51].

Although reaction-diffusion mechanisms like those described are indifferent to which side is left and which side is right, this decision evidently makes a difference biologically: While total inversion of the symmetry of internal organs (*situs inversus totalis*) usually has little adverse impact on health, partial inversions, in which the heart, for example, is predominantly on the right, are highly deleterious [63]. Perhaps for this reason vertebrate embryos have means to ensure that 99.99% of humans, for example, have standard left-right asymmetry.

In nonliving systems in which patterns self-organize by reaction-diffusion mechanisms (see, for example, Castets et al. [64] and Ouyang and Swinney [65]) the local initiators of activation typically arise by random fluctuations; patterns of spots and stripes of chemical concentration with generically similar appearance will form in such cases, but the detailed patterns will differ. In embryonic systems there is clearly a premium on having pattern forming mechanisms that generate the same results in each successive generation. In the axis-forming system of amphibians, birds and mammals just discussed, *monocilia* (motile extensions of the cell surface), at a localized site on the embryo midline (the Spemann organizer or primitive node), beat in an anticlockwise direction, creating fluid currents that bias the distribution of nodal and thus determine the direction of the broken symmetry [66]. The left-right asymmetry of the body follows from this early embryonic event that mobilizes a chemical-dynamic instability to produce a reliable morphological outcome.

5. Evolution of Developmental Mechanisms

Many key gene products that participate in and regulate multicellular development emerged over several billions of years of evolution in a world containing only single-celled organisms. Less than a billion years ago multicellular organisms appeared, and the gene products that had evolved in the earlier period were now raw material to be acted upon by physical and chemical-dynamic mechanisms on a more macroscopic scale [67]. The mechanisms described in the previous sections—chemical multistability,

chemical oscillation, and reaction-diffusion-based symmetry breaking, in conjunction with physical mechanisms such as changes in aggregate viscoelasticity and surface free energy, and adhesive differentials (both between cell types and across the surface of individual cells), would have caused these ancient multicellular aggregates to take on a wide range of biological forms. Any set of physicochemical activities that generated a new form in a reliable fashion within a genetically uniform population of cell aggregates would have constituted a primitive developmental mechanism.

But most modern-day embryos develop in a more rigidly programmed fashion: the operation of cell type- and pattern-generating physical and chemical-dynamic mechanisms is constrained and focused by hierarchical systems of coordinated gene activities—so-called “developmental programs.” These programs are the result of eons of molecular evolution that occurred mainly *after* the origination of multicellularity. The manner in which the morphological outcomes of physical and chemical-dynamic mechanisms may have been “captured” during early evolution to produce animal body plans, has been considered in earlier writings [7, 67, 68]. In the remainder of this chapter we will consider instead a model for how two different modes of segmentation in insects arose by variations in a common underlying chemical-dynamic mechanism.

5.1. Segmentation in insects

A major puzzle in the field of evolutionary developmental biology (“EvoDevo”) is the fact that evolutionarily-related organisms such as beetles (“short germ-band” insects) and fruit flies (“long germ-band” insects) have apparently different modes of segment formation. Similarly to somitogenesis in vertebrates (see Section 3), in short germ-band insects [69] (as well as in other arthropods, such as the horseshoe crab [70]), segmental primordia are added in sequence from a zone of cell proliferation (“growth zone”) (Fig. 2.9). In contrast, in long germ-band insects, such as the fruit fly *Drosophila*, a series of chemical stripes (i.e., parallel bands of high concentration of a molecule) forms in the embryo, which at this stage is a *syncytium*, a large cell with single cytoplasmic compartment containing about 6000 nuclei arranged in a single layer on the inner surface of the plasma membrane [71]. These stripes are actually alternating, evenly-spaced bands of transcription factors of the “pair-rule” class. The pair-rule genes include *even-skipped*, *fushi tarazu*, and *hairy*, which is the insect

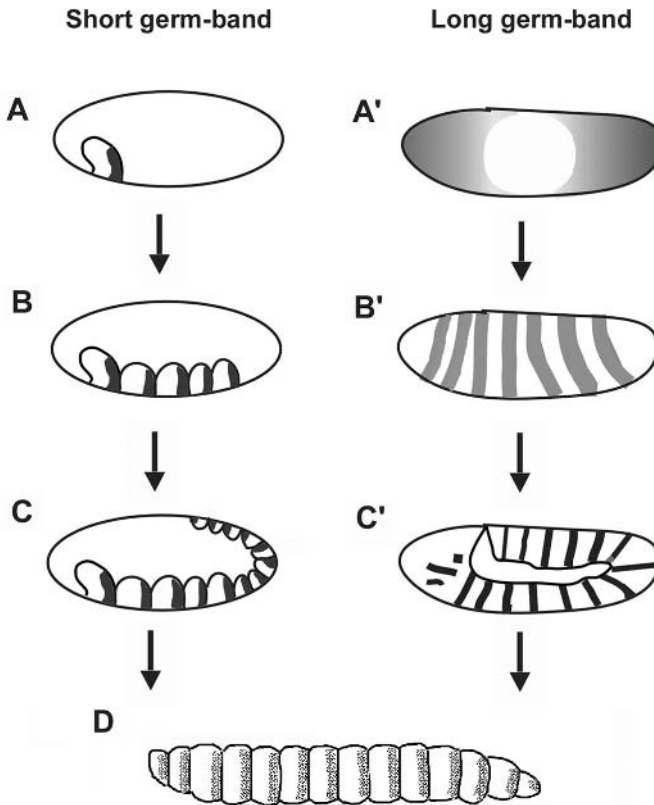


Figure 2.9. Schematic summary of segmentation modes in short germ-band and long germ-band insects. (Left, A) In short germ-band insects, one or groups of a few segments appear in succession. Black patches indicate expression of a segment polarity gene such as *engrailed*. (B) More segments appear posteriorly from a zone of proliferation. (C) The remainder of the segments form sequentially, as in B. (D) Idealized insect larva showing full array of segments. (Right, A') Long germ-band embryo with gradients of expression of maternal genes (e.g., *bicoid* and *nanos*) shown schematically. For simplicity, the patterns of gap gene expression (e.g., *hunchback*, *Krüppel*), intervening between steps A' and B' in *Drosophila*, are not shown. (B') Expression of pair-rule genes (e.g., *eve*, *ftz*, *hairy*) shown schematically in gray. (C') Expression of segment polarity genes (e.g., *engrailed*) shown in black. Adapted, with changes, from Salazar-Ciudad et al. (2001) [89].

homolog of the *c-hairy1* gene expressed in a periodic fashion during vertebrate somitogenesis (see Section 2). When cellularization (the enclosure of each nucleus and nearby cytoplasm in their own complete plasma membrane) takes place shortly thereafter, the cells of the resulting blastoderm will have periodically-distributed identities, determined by the particular mix of transcription factors they have incorporated. The different cell states are later transformed into states of differential adhesivity [72], and morphological segments form as a consequence.

No individual cell-based (“cell autonomous”) oscillations have thus far been identified during segmentation of invertebrates, unlike the case in vertebrates such as the mouse, chicken, and zebrafish. However, the sequential appearance of gene expression stripes from the posterior proliferative zone of short germ-band insects and other arthropods such as spiders, has led to the suggestion that these patterns in fact arise from a segmentation clock like that found to control vertebrate somitogenesis [73] (see Section 2.3).

On theoretical [74, 75] and experimental [76] grounds it has long been recognized that the kinetic properties that give rise to a chemical oscillation (such systems exhibit the “Hopf instability”; Section 3), can, when one or more of the components is diffusible, also give rise to standing or traveling spatial periodicities of chemical concentration (the “Turing instability”; Section 4). Considering embryonic tissues as excitable chemical-dynamic media can potentially unify the different segmentation mechanisms found in short and long germ-band insects. This would be quite straightforward if the *Drosophila* embryo were patterned by a reaction-diffusion system, which can readily give rise to a series of chemical standing waves (“stripes”).

In reality, however, *Drosophila* segmentation is controlled by a hierarchical system of genetic interactions that has little resemblance to the self-organizing pattern forming systems associated with reaction-diffusion coupling. The formation of overt segments in *Drosophila* (see St Johnston and Nusslein-Volhard [77] and Lawrence [71] for reviews) requires the prior expression of a stripe of the transcription factor product of the *engrailed* (*en*) gene in the cells of the posterior border of each of 14 presumptive segments [78]. The positions of the engrailed stripes are largely determined by the activity of the pair-rule genes *even-skipped* (*eve*) and *fuhs-i-tarazu* (*ftz*), which exhibit alternating, complementary seven stripe patterns prior to the formation of the blastoderm [79, 80].

However, despite the appearance of being produced by a reaction-diffusion system (Fig. 2.9), the stripe patterns of the pair-rule genes in *Drosophila* are generated rather by a complex set of interactions among transcription factors in the syncytium that encompasses the entire embryo. The formation of *eve* stripe number 2, for example, requires the existence of sequences in the *eve* promotor that switch on the *eve* gene in response to a set of spatially-distributed morphogens that under normal circumstances have the requisite values only at the stripe 2 position (Fig. 2.10) [81-83]. In particular, these promoter sequences respond to specific combinations of products of the “gap” genes (e.g., *giant*, *knirps*, the embryonically-produced version of *hunchback*). These proteins are transcription factors that are expressed in a spatially-nonuniform fashion and act as activators and competitive repressors of the pair-rule gene promoters [84] (also see discussion of the Keller model in Section 2). The patterned expression of the gap genes, in turn, is controlled by the responses of their own promoters to particular combinations of products of “maternal” genes (e.g., *bicoid*, *staufer*), which are distributed as gradients along the embryo at even earlier stages (Fig. 2.9). As the category name suggests, the maternal gene products are deposited in the egg during oogenesis.

While the expression of *engrailed* along the posterior margin of each developing segment seems to be a constant theme during development of arthropods, including those other than *Drosophila* (e.g., grasshoppers [69, 85], beetles [86]), the expression patterns of pair-rule genes is less well-conserved over evolution [87, 88]. The accepted view is that the short germ-band “sequential” mode is the more ancient way of making segments, and that the long germ-band “simultaneous” mode seen in *Drosophila*, which employs pair-rule stripes, is more recently evolved. The existence of “intermediate germ-band” insects, in which segmentation is sequential in one region of the embryo and simultaneous in another, suggests that *Drosophila* was derived from a short germ-band ancestor via such intermediate forms. Why and how cellularization of the blastula in some ancestral insects was impeded so as to produce a syncytium, is unknown.

5.2. Chemical dynamics and the evolution of insect segmentation

As noted above, the kinetic properties that give rise to a limit cycle chemical oscillation, can, when one or more of the components is diffusible, also

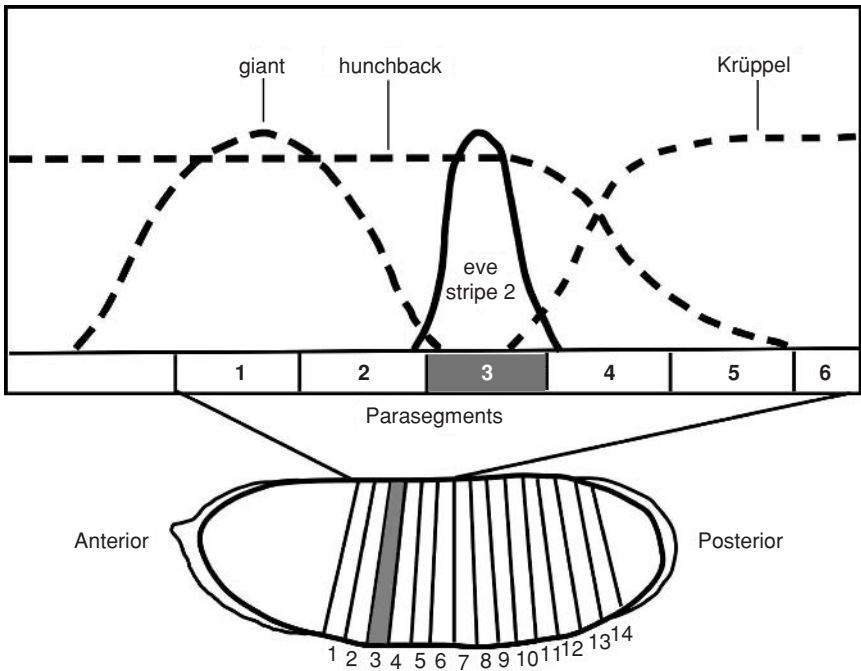


Figure 2.10. Specification of the second stripe of transcription of the *Drosophila even-skipped* gene. The *even-skipped* (*eve*) gene contains stripe-specific promoters responsive to different concentrations of the protein products of the gap-class genes. Stripe 2 of *eve* expression coincides with the third parasegment of the *Drosophila* embryo, the parasegments comprising the posterior region of one future morphological segment along with the anterior region of the future segment just posterior to it. Products of the gap genes, *giant*, *hunchback*, and *Krüppel* form gradients within the syncytium that activate *eve* expression at the stripe 2 location, while suppressing it to either side [81, 82]. Other *eve* stripes, expressed from the same gene, are specified by promoters responsive to different levels of gap gene products.

give rise to standing or travelling spatial periodicities of chemical concentration. Whether a system of this sort exhibits purely temporal, spatial, or spatiotemporal periodicity depends on particular ratios of reaction and diffusion coefficients. An important requirement of both these kinetic schemes is the presence of a direct or indirect positive autoregulatory circuit. A simple dynamical system that exhibits temporal oscillation or standing waves, depending on whether or not diffusion is permitted, is described by Salazar-Ciudad et al. [89] (Fig. 2.11).

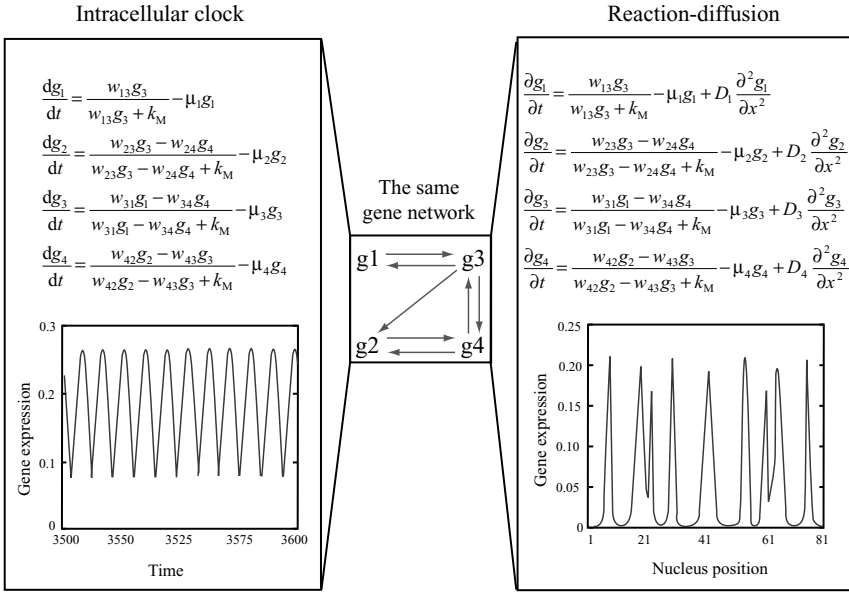


Figure 2.11. An example of a network that can produce (for the same parameter values) sequential stripes when acting as an intracellular biochemical clock in a cellularized blastoderm with a posterior proliferative zone, and simultaneously-forming stripes when acting in a diffusion-permissive syncytium. The network, based on the model of Salazar-Ciudad et al. (2001) [90] is shown in the central box. Black arrows indicate positive regulation and white arrows negative regulation. In the upper boxes, the equations governing each of the two behaviors are shown. The four genes involved in the central network diagram, as well as their levels of expression, are denoted by g_1 , g_2 , g_3 , and g_4 . In the reaction-diffusion case, g_1 and g_2 can diffuse between nuclei (note that the two set of equations differ only in the presence of a diffusion term for the products of genes 1 and 2). The lower boxes indicate the levels of expression of gene 2 for the two systems. For the intracellular clock the x -axis represents time, whereas in the reaction-diffusion system this axis represents space. In the 1 = D pattern shown, the initial condition consisted of all gene values set to zero except gene 1 in the central of 81 nuclei, which was assigned a small value (the exact value did not affect the pattern). The patterns shown were found when the following parameter values were used: $k_M = 0.01$; $w_{13} = 0.179$; $w_{23} = 0.716$; $w_{24} = -0.704$; $w_{31} = 0.551$; $w_{34} = -0.466$; $w_{42} = 0.831$; $w_{43} = -0.281$; $\mu_1 = 1.339$; $\mu_2 = 2.258$; $\mu_3 = 2.941$; $\mu_4 = 2.248$. For the reaction-diffusion case, the same parameter values are used but in addition: $D_1 = 0.656$ and $D_2 = 0.718$. From Salazar-Ciudad et al. [89], © Blackwell Publishing Co. Used with permission.

Based on the experimental findings, theoretical considerations and evolutionary inferences described above, it was hypothesized that the ancestor of *Drosophila* generated its segments by a reaction-diffusion system [89, 91], built upon the presumed chemical oscillator underlying short germ-band segmentation. Modern-day *Drosophila* contains (retains, according to the hypothesis) the ingredients for this type of mechanism. Specifically, in the syncytial embryo several of the pair-rule proteins (e.g., *eve*, *ftz*) diffuse over short distances among the cell nuclei that synthesize their mRNAs, and positively regulate their own synthesis [92-94].

The hypothesized evolutionary scenario can be summarized as follows: the appearance of the syncytial mode of embryogenesis converted the chemical oscillation-dependent temporal mechanism found in the more ancient short germ-band insects into the spatial standing wave mechanism seen in the more recently evolved long germ-band forms. Of course, the pair-rule stripes formed by the proposed reaction-diffusion mechanism would have been equivalent to one another. That is, they would have been generated by a single mechanism acting in a spatially-periodic fashion, not by stripe-specific molecular machinery (see above). Thus despite suggesting an underlying physical connection between modern short germ-band segmentation and segmentation in the presumed ancestor of long germ-band forms, this hypothesis introduces a new puzzle of its own: Why does modern-day *Drosophila* not use a reaction-diffusion mechanism to produce its segments?

5.3. Evolution of developmental robustness

Genes are always undergoing random mutation, but morphological change does not always track genetic change. Particularly interesting are those cases in which the outward form of a body plan or organ does not change, but its genetic “underpinning” does. This can result from particular kind of natural selection, termed “canalizing selection” by Waddington [95] (see also Schmalhausen [96]). Canalizing selection will preserve those randomly acquired genetic alterations that happen to enhance the reliability of a developmental process. Development would thereby become more complex at the molecular level, but correspondingly more resistant (“robust”) to external perturbations or internal noise that could disrupt non-reinforced physical mechanisms of determination. Indeed, the patterns formed by reaction-diffusion systems are notoriously sensitive to

temperature and domain size [97, 98], and any developmental process that was solely dependent on this mechanism would be unreliable.

If the striped expression of pair-rule genes in the ancestor of modern *Drosophila* was generated by a reaction-diffusion mechanism, this inherently variable developmental system would have been a prime candidate for canalizing evolution. The elaborate systems of multiple promoter elements responsive to pre-existing, nonuniformly distributed molecular cues (e.g., maternal and gap gene products), seen in *Drosophila* is therefore not inconsistent with this pattern having originated as a reaction-diffusion process.

The question was raised above as to why modern-day *Drosophila* does not use a reaction-diffusion mechanism to produce its segments. In light of the discussion in the previous paragraphs, we can consider the following tentative answer: pattern-forming systems based on reaction-diffusion may be inherently unstable evolutionarily (for the same reasons of sensitivity to parameters and domain size that make them dynamically unstable), and would therefore have been replaced, or at least reinforced, by more hierarchically-organized genetic control systems under pressure of natural selection. A corollary of this hypothesis is that the patterns produced by modern, highly evolved, pattern-forming systems would be robust in the face of further genetic change.

Similar canalizing evolution may also have occurred in short-germ band insects—not as much is known about the underlying mechanism of segmentation in these groups. However, it is also possible that oscillatory mechanisms (which continue to be used in vertebrates [23], and may be used in short germ-band insects) are less easy to replace by more reliable alternatives (i.e., hierarchies of genetic regulation) than are standing wave mechanisms. (This issue is discussed in greater detail in Salazar-Ciudad et al. [89].)

The evolutionary hypothesis of self-organization followed by canalization has been examined computationally in a simple physical model by Salazar-Ciudad and coworkers [90]. The model consists of N_c nuclei arranged in a row within a syncytium. Each nucleus has the same genome (i.e., the same set of N_g genes), and the same *epigenetic system* (i.e., the same activating and inhibitory relationships among these genes). Genes in these networks interact according to a set of simple rules that embody unidirectional interactions in which an upstream gene activates a downstream one, as well as reciprocal interactions, in which genes feed back

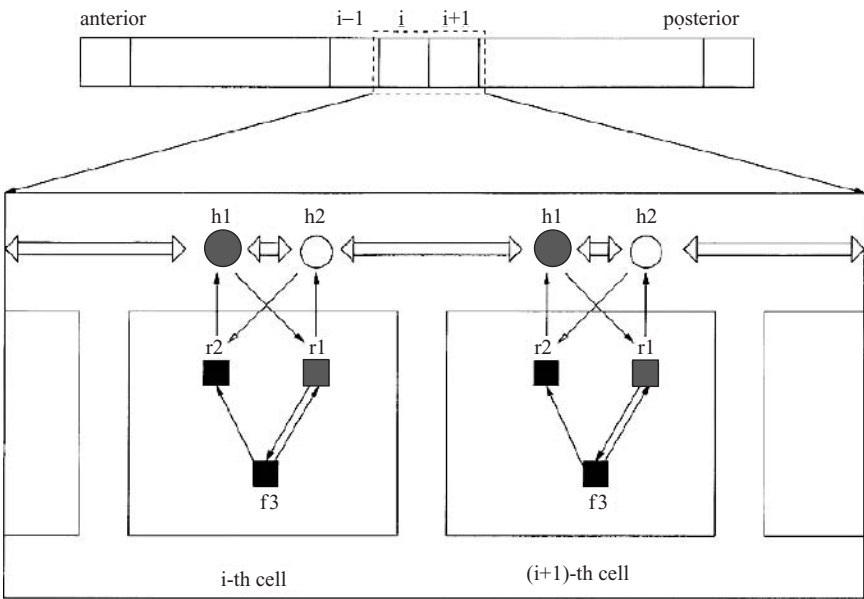


Figure 2.12. Schematic representation of the kinds of interactions included in the gene network evolution model of Salazar-Ciudad et al. The boxes denote genes acting inside cells whereas circles denote diffusible paracrine factors. Abbreviations: h designates a hormone or paracrine factor, r a receptor, and f a transcription factor. Arrows with black arrowheads indicate positive interactions and arrows with white arrowheads indicate inhibitory interactions. Thick horizontal arrows indicate diffusion. From Salazar-Ciudad et al. (2001) [90]. Used with permission.

(via their products) on each other's activities (Fig. 2.12). This formalism was based on a similar one devised by Reinitz and Sharp [99], who considered the specific problem of segmentation in the *Drosophila* embryo. Gene products interact by binding to or modifying one another, or by binding to cis-acting (i.e., located on the same DNA strand as the gene) promoter DNA sequences, as described above in the Keller model [12]. A subset of the genes (N_p) specify diffusible (paracrine) factors. The model assumes that the change in gene expression induced by an interaction follows a saturating Hill function (a class of function widely used for molecular binding processes [100]). Promoters are characterized by the values of coupling constants W_{jk} that weight both the affinity of the transcriptional factor for the promoter and the intensity of the response produced by the binding.

The gene regulatory networks considered by Salazar-Ciudad and coworkers are modeled by a dynamic system obeying the following set of equations, which describe the change in the concentration in each of the N_c nuclei (numbered by the first index) of each of the N_g gene product (numbered by the second index), due to the effect of all other genes:

Non-diffusible gene products:

$$\frac{\partial g_{ij}}{\partial t} = \frac{\Phi(h_{ij})}{k_M + \Phi(h_{ij})} - \mu_j g_{ij}, \quad h_{ij} = \sum_{k=1}^{N_g} W_{jk} g_{jk} \quad (5.1)$$

$$(i = 1, 2, \dots, N_c, j = N_p + 1, \dots, N_g)$$

Diffusible gene products:

$$\frac{\partial g_{il}}{\partial t} = \frac{\Phi(h_{il})}{k_M + \Phi(h_{il})} - \mu_l g_{il} + D_l \nabla^2 g_{il}, \quad h_{il} = \sum_{k=1}^{N_g} W_{lk} g_{lk} \quad (5.2)$$

$$(i = 1, 2, \dots, N_c, l = 1, \dots, N_p)$$

The Heaviside function, $\Phi(\Phi(x) = x$ for $x > 0$ and $\Phi(x) = 0$ otherwise), ensures that inhibiting interactions do not lead to negative concentration values. D_l and μ_l are, respectively, the diffusion coefficient and the intrinsic rate of degradation for gene product g_l , and k_M is a Michaelis-type constant defining the rate of response to an activator or inhibitor.

One may ask whether a system of this sort, with particular values of the gene-gene coupling constants W_{jk} , can form a pattern. Salazar-Ciudad and coworkers performed simulations on systems containing 25 nuclei [90]. Zero-flux boundary conditions were used (i.e., the boundaries of the domain were impermeable to diffusible morphogens), and initial conditions were set such that at $t = 0$ the levels of all gene products had zero value except for that of an arbitrarily chosen gene, which had a non-zero value in the nucleus at the middle position. A pattern was considered to arise if after some time different nuclei stably expressed one or more of the genes at different levels. Isolated single nuclei, or isolated patches of contiguous nuclei expressing a given gene, are the 1-dimensional analogue of isolated stripes of gene expression in a 2-dimensional sheet of nuclei, such as that in the *Drosophila* embryo prior to cellularization [90].

It had earlier been determined that the core mechanisms responsible for stable patterns fell into two non-overlapping topological categories [101]. Mechanisms that form patterns by virtue of the unidirectional influence

of one gene on the next, in an ordered succession, as in the “maternal gene induces gap gene induces pair-rule gene” scheme described above for aspects of *Drosophila* segmentation, were termed “hierarchical.” In contrast, mechanisms in which reciprocal positive and negative feedback interactions give rise to the pattern, were termed “emergent.” Emergent systems are equivalent to dynamical systems, such as the transcription factor networks discussed in Section 2 and the reaction-diffusion systems discussed in Section 2.4.

This is not to imply that an entire embryonic patterning process—the formation of a pair-rule stripe in *Drosophila*, for example—is wholly hierarchical or emergent. Rather, it indicates that the process can be decomposed into modules that are unambiguously of one or the other topology [101].

In their computational studies of the evolution of developmental mechanisms Salazar-Ciudad and co-workers identified emergent and hierarchical networks that produced a particular pattern—e.g., three “stripes” [90]. (Note that patterns themselves are neither emergent or hierarchical—these terms apply to the mechanisms that generate them.) They asked if given networks would “breed true” phenotypically, despite changes to their underlying circuitry. That is, would their genetically altered “progeny” exhibit the same pattern as the unaltered version? Genetic alterations in these model systems consisted of point mutations (i.e., changes in a W_{jk} value), duplications, recombinations (i.e., interchange of various W_{jk} values between genes) and the acquisition of new interactions (i.e., a W_{jk} that was initially equal to zero was randomly assigned a small positive or negative value).

To evaluate the consequence of such alterations it was necessary to define a metric of “distance” between different patterns. Roughly, this was done by specifying the state of each nucleus in a model syncytium in terms of the value of the gene product forming a pattern. Two patterns were considered to be equivalent if all nuclei in corresponding positions were in the same state. The degree of divergence from such equivalence could then be quantified [90].

It was found that emergent networks were much more likely to diverge from the original pattern than hierarchical networks after undergoing such simulated evolution (Fig. 2.13). In other simulations the pattern itself was held constant (i.e., only those genetic variants that had the same number of “stripes” as the original were retained after each iteration) and it was found that networks that started out as emergent could be converted into hierarchical networks.

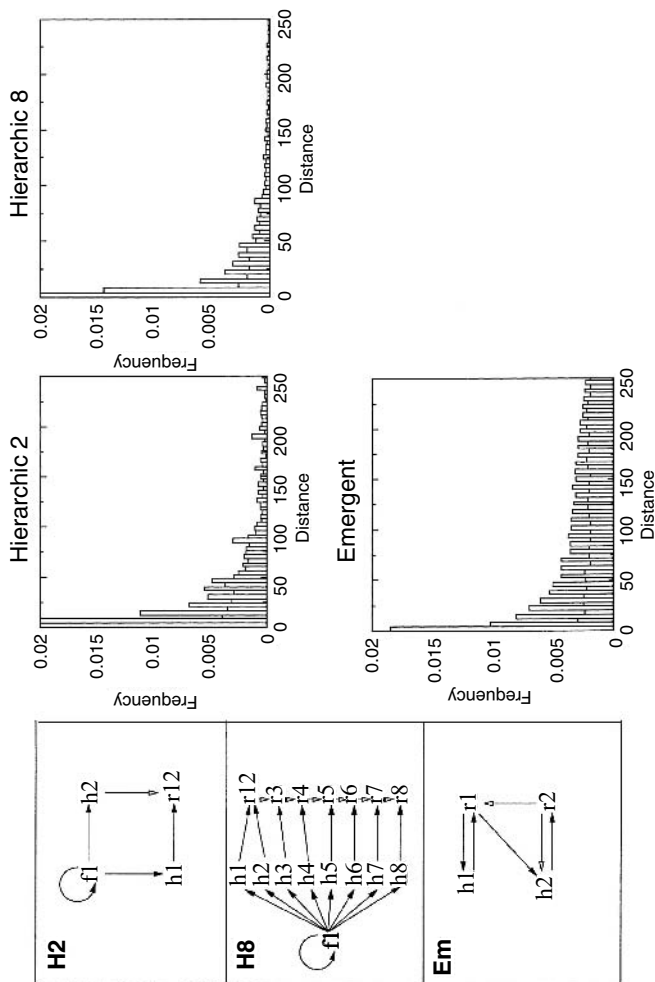


Figure 2.13. (left panels) Three network topologies studied in simulated evolution experiments by Salazar-Ciudad et al. [90]. The terms h , r and f , and arrows are as defined in the legend to Figure 2.12. “Hierarchic 2” (H2) and “Hierarchic 8” (H8) refer to hierarchical network topologies with two paracrine factors and one receptor, and eight paracrine factors and seven receptors, respectively. “Emergent” (Em) is a network topology with two paracrine factors and two receptors. (right panels) Estimation of the distance between genotype and phenotype in the three networks topologies shown on the left. The x axis represents the distance between a network and a mutated version. (See text and original article [90] for definition of distance measure.) The y axis represents the relative frequency of such distances between the 100,000 networks of each topology analyzed and 30 randomly mutated versions of each. Adapted, with changes, from Salazar-Ciudad et al. (2001) [90]. Used with permission.

Subject to the caveats to what is obviously a highly schematic analysis, the implications of these computational experiments for the evolution of segmentation in long germ-band insects are the following: (i) if the ancestral embryo indeed generated its seven-stripe pair-rule protein patterns by a reaction-diffusion mechanism, and (ii) if this pattern was sufficiently well-adapted to its environment so as to provide a premium on breeding true, then (iii) genetic changes that preserved the pattern but converted the underlying network from an emergent one to a hierarchical one (as seen in present-day *Drosophila*) would have been favored.

6. Conclusions

Biological development depends on genes and their products, and biological evolution depends on genetic change. But it is a mistake to consider development as being the result of programmed gene expression changes alone, or to consider evolution to be caused by genetic variation and selection exclusively. The cell types and tissue patterns and forms that arise during development are manifestations of complex systems, and system behaviors are not simply a matter of inventories of components and their interconnections. Moreover, the changes in form and function induced by genetic mutation are also determined by the system properties of the organism, particularly those in play when it is taking form, i.e., during development. Therefore, even though genetic mutation may be random and natural selection opportunistic, evolutionary change has preferred directions.

None of this will be surprising to physical scientists who study complex systems. Despite earlier precedents, (e.g., Rashevsky [102], Bertalanffy [103, 104]), that were before their time in that the molecular components of the cell physiological and developmental systems discussed were unknown, the systems way of thinking is only now becoming part of the biological mainstream, though there have been notable landmarks along the way (e.g., Goodwin [10], Winfree [105], Meinhardt [97]).

The models we have presented in the preceding sections are all unified in that they deal with chemical-dynamical aspects of multicellular development and evolution. Other physical aspects of the same questions are reviewed elsewhere [106-109]. The examples considered in this chapter represent several different aspects of the new systems biology. The most

abstract of the models are those dealing with cell differentiation. While the autoregulatory transcription factor network model of Keller [12] deals with general categories of gene-gene interactions rather than specifically-named molecules, all the categories considered, including the one described in detail in Section 2.2, are based on molecular biological findings. Multistability is essentially a direct consequence of these interactions. The representation of cells and molecules in the Kaneko-Yomo isologous diversification model is more distant from any experimental data, but does not contradict anything known about the formal properties of such systems. The unique thing about this model is that as an exercise in dynamical systems theory motivated by biological questions, it has actually uncovered a dynamical phenomenon—interaction-dependent attractors—not previously appreciated. If this phenomenon indeed proves relevant to developmental properties such as the community effect [15], it would be a rare case (Turing’s 1952 description of the reaction-diffusion instability [52] is another) where biologically-motivated mathematics led to a fundamental finding as well as a solution to a biological problem.

As hypothesized by Bateson [21] and Cooke and Zeeman [26], cell autonomous chemical oscillations are clearly relevant to vertebrate segmentation [22]. The biochemical oscillator analysis of Lewis [27] and Monk [28] is based on highly detailed knowledge of the molecular and cellular interactions occurring during zebrafish somitogenesis. The introduction of a time-lag, a key component in causing the oscillator to behave realistically, is a mathematical maneuver that is virtually dictated by the molecular biology of gene expression [27]. Along with Meinhardt’s molecularly-cognizant analysis of the amphibian axis-forming system as excitations of an excitable medium [49], the analyses of Lewis, as well as the important work, not discussed here, of Odell and coworkers [110, 111] on the chemical dynamics of developmental modules, represent the leading edge of an integrated experimental-mathematical approach to systems phenomena during development.

The models of Lewis and Meinhardt, while having the formal structure of chemical-dynamic schemes, make use of chemically nonexplicit (“black box”) cell response functions to describe the synthesis and breakdown of their gene product components. For this reason, Hentschel et al. [49] who, like Meinhardt [49], use a Turing mechanism-inspired approach to model an aspect of embryogenesis (in their case the skeletal pattern of the vertebrate limb), propose that the term “reactor-diffusion” be

used instead of reaction-diffusion to describe these hybrid cell-chemical systems.

In analyzing the complex question of the evolution of a developmental mechanism, Salazar-Ciudad et al. [89, 90] resort to schematic simplifications analogous to those used by Keller [12] and Kaneko and coworkers [17-20]. Like those models, that of Salazar-Ciudad and coworkers employs gene-gene product interactions consistent with experimental knowledge of the relevant systems, in this case a simplified version of a molecularly-based model for the formation of pair-rule stripes in *Drosophila* [99]. The evolutionary simulations performed with this model suggest a plausible scenario for the increasingly-recognized fact that developmental mechanisms may evolve, even when the morphologies they generate remain static over vast periods [7]. In fact, canalizing selection can reasonably have transformed an embryonic pattern that may have originated as a set of reactor-diffusion processes analogous to the chemical-dynamic mechanisms discussed earlier in this chapter (i.e., the pair-rule stripes), into the hierarchical genetic program seen in the modern fruit-fly.

Sections 2.2–2.4 of this chapter indicated the operation of chemical-dynamic processes in differentiation, segmentation and axis formation of modern-day organisms. Nonetheless, the enormous degree of apparently “hard-wired” molecular integration seen in some developmental systems has been used in the past to support the premise that these physicochemical determinants are of only marginal relevance. The discussion in Section 2.5 suggests that by taking evolution into account, we can reconcile developmental systems with the dynamics of the less constrained systems from which they originated.

Acknowledgment

We acknowledge Nation Science Foundation grant No. IBN - 0083653 under the Bio-complexity Initiative, for support.

References

1. S. F. Gilbert, M. S. Tyler, and R. N. Kozlowski, *Developmental Biology*, 7th ed., Sinauer Associates, Sunderland, MA (2003).
2. S. H. Strogatz, *Nonlinear Dynamics and Chaos With Applications to Physics, Biology, Chemistry, and Engineering*. Perseus, Cambridge, MA (1994).
3. International Human Genome Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

4. M. J. Lercher, A. O. Urrutia, and L. D. Hurst, *Nat Genet* **31**, 180-183 (2002).
5. M. T. Borisuk, J. J. Tyson, *J Theor Biol* **195**, 69-85 (1998).
6. M. B. Elowitz and S. Leibler, *Nature* **403**, 335-338 (2000).
7. S. A. Newman and G. B. Müller, *J Exp Zool* **288**, 304-317 (2000).
8. E. Jablonka and M. J. Lamb, *Epigenetic Inheritance and Evolution*. Oxford University Press, Oxford, U.K. (1995).
9. S. A. Kauffman, *J Theor Biol* **22**, 437-467 (1969).
10. BC Goodwin, *Temporal Organization in Cells; A Dynamic Theory of Cellular Control Processes*. Academic Press, London, New York (1963).
11. M. Mannervik, Y. Nibu, H. Zhang, and M. Levine *Science* **284**, 606-609 (1999).
12. A. D. Keller, *J Theor Biol* **172**, 169-185 (1995).
13. F. G. Giancotti and E. Ruoslahti, *Science* **285**, 1028-1032 (1999).
14. C. Morisco, K. Seta, S. E. Hardt, Y. Lee, S. F. Vatner, and J. Sadoshima, *J Biol Chem* **276**, 28586-28597 (2001).
15. J. B. Gurdon, *Nature* **336**, 772-774 (1988).
16. H. J. Standley, A. M. Zorn, and J. B. Gurdon, *Development* **128**, 1347-1357 (2001).
17. K. Kaneko and T. Yomo, *Physica D* **75**, 89-102 (1994).
18. K. Kaneko and T. Yomo, *Bull Math Biol* **59**, 139-196 (1997).
19. C. Furusawa and K. Kaneko, *Bull Math Biol* **60**, 659-687 (1998).
20. K. Kaneko, Organization Through Intra-Inter Dynamics, in *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*, G. B. Müller and S. A. Newman (eds.), MIT Press, Cambridge, MA (2003).
21. W. Bateson, *Materials for the Study of Variation*. London, Macmillan (1894).
22. I. Palmeirim, D. Henrique, D. Ish-Horowicz, and O. Pourquie, *Cell* **91**, 639-48 (1997).
23. O. Pourquie, A biochemical oscillator linked to vertebrate segmentation, in *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*. Müller G. B. and Newman S. A. (eds.), Cambridge, MA, MIT Press (2003).
24. J. Dubrulle, M. J. McGrew, and O. Pourquie, *Cell* **106**, 219-232 (2001).
25. O. Pourquie, *Science* **301**, 328-330 (2003).
26. J. Cooke and E. C. Zeeman, *J Theor Biol* **58**, 455-476 (1976).
27. J. Lewis, *Curr Biol* **13**, 1398-1408 (2003).
28. N. A. Monk, *Curr Biol* **13**, 1409-1413 (2003).
29. S. A. Holley, R. Geisler, and C. Nusslein-Volhard, *Genes Dev* **14**, 1678-1690 (2000).
30. A. C. Oates and R. K. Ho, *Development* **129**, 2929-2946 (2002).
31. C. Takke and J. A. Campos-Ortega, *Development* **126**, 3005-3014 (1999).
32. S. A. Holley, D. Julich, G. J. Rauch, R. Geisler, and C. Nusslein-Volhard, *Development* **129**, 1175-1183 (2002).
33. S. Artavanis-Tsakonas, M. D. Rand, and R. J. Lake, *Science* **284**, 770-776 (1999).
34. C. D. Stern and R. Bellairs, *Anat Embryol (Berl)* **169**, 97-102 (1984).

35. D. R. Primm, W. E. Norris, G. J. Carlson, R. J. Keynes, and C. D. Stern *Development* **105**, 119-130 (1989).
36. P. D. Nieuwkoop, *Wilhelm Roux' Arch. Entw. Mech. Org.* **162**, 341-373 (1969).
37. J. Green, *Dev Dyn* **225**, 392-408 (2002).
38. N. McDowell, J. B. Gurdon, and D. J. Grainger, *Int J Dev Biol* **45**, 199-207 (2001).
39. B. Christen and J. Slack, *Development* **126**, 119-125 (1999).
40. M. Kerszberg and L. Wolpert, *J Theor Biol* **191**, 103-114 (1998).
41. A. D. Lander, Q. Nie, and F. Y. *Dev Cell* **2**, 785-796 (2002).
42. E. V. Entchev, A. Schwabedissen, and M. Gonzalez-Gaitan, *Cell* **103**, 981-991, (2000).
43. A. A. Teleman and S. M. Cohen, *Cell* **103**, 971-980 (2000).
44. H. F. Nijhout, *The Development and Evolution of Butterfly Wing Patterns*. Smithsonian Institution Press, Washington (1991).
45. S. Kondo and R. Asai, *Nature* **376**, 765-768 (1995).
46. T. Jiang, H. Jung, R. B. Widellitz, and C. Chuong, *Development* **126**, 4997-5009 (1999).
47. S. A. Newman and H. L. Frisch, *Science* **205**, 662-668 (1979).
48. H. G. E. Hentschel, T. Glimm, J. A. Glazier, and S. A. Newman, *Proc Roy Soc London B Biol Sci* **271**, 1713-1722 (2004).
49. H. Meinhardt, *Int J Dev Biol* **45**, 177-188 (2001).
50. H. Spemann and H. Mangold, *Wilhelm Roux' Arch. Entw. Mech. Org.* **100**, 599-638 (1924).
51. L. Solnica-Krezel, *Curr Biol* **13**, R7-9 (2003).
52. A. Turing, *Phil Tran. Roy Soc Lonon B* **237**, 37-72 (1952).
53. E. Van Obberghen-Schilling, N. S. Roche, K. C. Flanders, M. B. Sporn, and A. Roberts, *J. Biol. Chem.* **263**, 7741-7746 (1988).
54. B. Sun, S. Bush, L. Collins-Racie, E. LaVallie, E. DiBlasio-Smith, N. Wolfman, J. McCoy, and H. Sive, *Development* **126**, 1467-1482 (1999).
55. D. Clements, R. V. Friday, and H. R. Woodland, *Development* **126**, 4903-4911 (1999).
56. S. Schulte-Merker and J. C. Smith, *Curr Biol* **5**, 62-67 (1995).
57. R. Harland and J. Gerhart, *Annu. Rev. Cell. Dev. Biol.* **13**, 611-667 (1997).
58. H. Juan and H. Hamada, *Genes Cells* **6**, 923-930 (2001).
59. W. W. Branford and HJ Yost, Lefty, *Curr. Biol.* **12**, 2136-2141 (2002).
60. M. Yamamoto, Y. Saijoh, A. Perea-Gomez, W. Shawlot, R. R. Behringer, S. L. Ang, H. Hamada, and C. Meno, *Nature* **428**, 387-392 (2004).
61. Y. Chen and A. F. Schier, *Curr. Biol.* **12**, 2124-2128 (2002).
62. R. Sakuma, Y. Y. Ohnishi, C. Meno, H. Fujii, H. Juan, J. Takeuchi, T. Ogura, E. Li, K. Miyazono, and H. Hamada, *Genes Cells* **7**, 401-412 (2002).
63. A. S. Aylsworth, *Am. J. Med. Genet.* **101**, 345-355 (2001).
64. V. Castets, E. Dulos, J. Boissonade, and P. DeKepper *Phys. Rev. Lett.* **64**, 2953-2956 (1990).
65. Q. Ouyang and H. Swinney, *Nature*, **352**, 610-612 (1991).

66. S. Nonaka, H. Shiratori, Y. Saijoh, and H. Hamada, *Nature* **418**, 96-99 (2002).
67. S. A. Newman, From Physics To Development: The Evolution of Morphogenetic Mechanisms, in *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*. GB Müller and SA Newman (eds.), Cambridge, MA, MIT Press (2003).
68. S. A. Newman, *J. Ev. Biol.* **7**, 467-488 (1994).
69. N. H. Patel, T. B. Kornberg, and C. S. Goodman, *Development* **107**, 201-212 (1989).
70. T. Itow, *Roux's Arch. Dev. Biol* **195**, 323-333 (1986).
71. P. A. Lawrence, *The Making of a fly: The Genetics of Animal Design*, Blackwell Scientific, Oxford, Boston (1992).
72. K. D. Irvine and E. Wieschaus, *Development* **120**, 827-841 (1994).
73. A. Stollewerk, M. Schoppmeier, and W. G. Damen, *Nature* **423**, 863-865 (2003).
74. J. Boissonade, E. Dulos, and P. DeKepper, Turing Patterns: From Myth to Reality, in *Chemical Waves and Patterns*. R Kapral and K Showalter (eds.), Kluwer, Boston (1994).
75. C. B. Muratov, *Phys. Rev.* **E55**, 1463-1477 (1997).
76. I. Lengyel and I. R. Epstein, *Proc. Natl. Acad. Sci. USA* **89**, 3977-3979 (1992).
77. D. St Johnston and C. Nusslein-Volhard, *Cell* **68**, 201-219 (1992).
78. T. L. Karr, M. P. Weir, Z. Ali, and T. Kornberg, *Development* **105**, 605-612 (1989).
79. M. Frasch and M. Levine, *Genes. Dev.* **1**, 981-995 (1987).
80. K. Howard and P. Ingham, *Cell* **44**, 949-957 (1986).
81. S. Small, A. Blair, and M. Levine, *EMBO J.* **11**, 4047-4057 (1992).
82. S. Small, R. Kraut, T. Hoey, R. Warrior, and M. Levine, *Genes. Dev.* **5**, 827-839 (1991).
83. S. Small, A. Blair, and M. Levine, *Dev. Biol.* **175**, 314-324 (1996).
84. D. E. Clyde, M. S. Corado, X. Wu, A. Pare, D. Papatsenko, and S. Small, *Nature* **426**, 849-853, (2003).
85. N. H. Patel, EE Ball, and CS Goodman, *Nature* **357**, 339-342 (1992).
86. S. J. Brown, N. H. Patel, and R. E. Denell, *Dev. Genet.* **15**, 7-18 (1994).
87. R. Dawes, I. Dawson, F. Falciani, G. Tear, and M. Akam, *Development* **120**, 1561-1572, (1994).
88. S. J. Brown, R. B. Hilgenfeld, and R. E. Denell, *Proc. Natl. Acad. Sci. USA* **91**, 12922-12926, (1994).
89. I. Salazar-Ciudad, R. Solé, and S. A. Newman, *Evolution & Development* **3**, 95-103 (2001).
90. I. Salazar-Ciudad, S. A. Newman, and R. Solé, *Evolution & Development* **3**, 84-94 (2001).
91. S. A. Newman, *BioEssays* **15**, 277-283 (1993).
92. K. Harding, T. Hoey, R. Warrior, and M. Levine, *Embo. J.* **8**, 1205-1212 (1989).
93. D. Ish-Horowicz, S. M. Pinchin, P. W. Ingham, and H. G. Gyrkovics, *Cell* **57**, 223-232 (1989).

94. A. F. Schier and W. J. Gehring, *Embo. J.* **12**, 1111-1119 (1993).
95. C. H. Waddington, *The Strategy of the Genes*. Allen and Unwin, London (1957).
96. I. I. Schmalhausen, *Factors of Evolution*. Blackstone, Philadelphia (1949).
97. H. Meinhardt, *Models of Biological Pattern Formation*, Academic, New York, (1982).
98. L. J. Harrison, *Kinetic Theory of Living Form*. Cambridge, Cambridge University Press (1993).
99. J. Reinitz, E. Mjolsness, and D. H. Sharp, *J. Exp. Zool.* **271**, 47-56 (1995).
100. A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, revised edition, Ashgate, London, Brookfield, VT (1995).
101. I. Salazar-Ciudad, J. Garcia-Fernandez, and R. V. Sole, *J. Theor. Biol.* **205**, 587-603 (2000).
102. N. Rashevsky, *Mathematical Biophysics*, revised edition, Univ. of Chicago Press, Chicago (1948).
103. L. V. Bertalanffy, *Vom Molekül zur Organismenwelt; Grundfragen der Modernen Biologie*, 2. verb. Aufl. ed. Akademische Verlagsgesellschaft Athenaion, Potsdam (1949).
104. L. V. Bertalanffy: *General System Theory; Foundations, Development, Applications*, G. Braziller, New York, (1969).
105. A. T. Winfree, *The Geometry of Biological Time*. Springer-Verlag, New York, (1980).
106. S. A. Newman and WD Comper, *Development* **110**, 1-18, (1990).
107. G. B. Müller and SA Newman, *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*. MIT Press, Cambridge, MA, (2003).
108. H. Levine and E. Ben-Jacob, *Phys. Biol.* **1**, 2004, in press,.
109. G. Forgacs and S. A. Newman, *Biological Physics of the Developing Embryo*. Cambridge Univ. Press, Cambridge, MA, in press.
110. G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, *Nature* **406**, 188-192 (2000).
111. E. Meir, G. von Dassow, E. Munro, and G. M. Odell, *Curr. Biol.* **12**, 778-786 (2002).

Chapter 3

THE CIRCLE THAT NEVER ENDS: CAN COMPLEXITY BE MADE SIMPLE?

Donald C. Mikulecky

Professor Emeritus and Senior Fellow in the Center for Biological Complexity
Virginia Commonwealth University, Richmond, Virginia

1. Introduction: The Nature of the Problem and Why it Has No Clear Solution

There is no need to try to ease into the discussion because the issue being discussed here is too involved to waste words that way. It is important to get right to the point. The starting premise is that we seek to understand the world in which we live. This leads immediately to a whole set of related premises that define the problem as being the answer to the question “how we can obtain that which we seek.” It is really the question “why is it so hard to understand the world around us” that should be asked, but that is the end of the story, not the beginning. The initial premise also leads to another, related, set of premises involving the existence of that world and our ability to ever know it completely, whatever “completely” might mean in this context. This is already a large problem. Here is why. As this is being written a number of things are happening. One possibility is that the author is merely stating the obvious. If that is true then no one should have reason to question what is being written. Another possibility is that the author is treading on at least some new ground. If that is true, then it is also possible that some of this new ground contributes to the body of knowledge we have already produced about the world we seek to understand. There is the problem. If the reader did not balk at the notion that the meaning of “completely” depends on a context of some kind then the problem should be clear. The context has to either be frozen at some state of history or it changes as we go along. One might argue that the changes are so slight as to be inconsequential for a discussion of this limited duration. That argument skirts around the issue. The issue is the circularity of context dependence.

Classical science and its underlying philosophy have tried very hard to eliminate this context dependency, yet these very intensely focused efforts by the best minds we have had have merely made it clearer and clearer that the problem is here to stay and must be dealt with.

It would be very pretentious to claim that large insights into how to deal with this issue are going to be presented in a review of this duration. Rather, some glimpses at the nature of the situation as this author sees it are all that can be provided. Those glimpses have a history and the author has strong prejudices about where we are in our attempts to deal with this matter. In the spirit of Karl Popper's ideas (Popper [1], Dress [2]) about the necessary subjectivity of any scientist/scholar, those prejudices need to be made clear from the onset. For that reason, what follows will be based on one person's career and that person's attempts to understand the world, in keeping with the original premise. It is important to acknowledge the roles of a number of scientific leaders in the forming of this worldview. Julian Tobias, as a neurophysiologist and as a Ph. D. thesis mentor, asked many important questions that could not be answered. Aaron Katchalsky [3] presented an attitude and approach to solving the problem of how we can hope to understand the world that was powerful and unique. His death at the hands of terrorists in June of 1972 changed both scientific and world history, in the opinion of this author. Leonardo Peusner [4-12] as a graduate student in the Harvard Biophysics Lab helped by the work of Katchalsky and his students George Oster and Alan Perelson [13] showed an extended structure to our scientific model of the physical world. Later, Eric Schneider and James Kay [14] wove this together with even more to provide a picture of the ecosystems that are entwined as life on this planet. Then early on and again and again there were the ideas of Robert Rosen. (Rosen [15-20], Mikulecky [21]). He died not that many years ago and his family was kind enough to allow me to have some copies of his unpublished work. His ideas made it clear that there has to be a new way of looking at the *process* of knowing. That is the approach to the problem to be briefly described here. One more thing needs to be dealt with before going any further. Is this science? Is it philosophy? In the world of knowledge that can be compartmentalized they each have their place. These are the kind of questions that can only be destructive in this context. The reason should be clear after the exposition has been ended. For now it will be necessary to accept another premise: *The nature of the world out there is such that the idea that much is lost by trying to reduce it to parts is paramount. The whole*

is always more than, and often different from, the sum of its parts. This is true whether we are talking about the material world or our thoughts about that world. Anyone who has trouble accepting this premise will probably find what follows difficult to accept. Please give the entire development a chance before dismissing it. That is one distinct advantage of the brevity of the presentation.

1.1. The human mind and the external world

What is the connection between the thoughts going on in our minds at the moment and the existence of an “objective” and “physical” world in which that mind somehow has its existence? The nature of that connection is the key to everything that follows. There would not be a notion of an external world were it not for the constant input of sensory “signals” for lack of a better word. Sensory physiology is a fascinating subject. It deals with the way these signals are able to impinge on specialized “receptors” and undergo a transduction into nerve impulses that are “all or none”. The all or none concept is the finding that nerves send signals of a given magnitude and they either fire or they don’t. The intensity or strength of a stimulus is encoded by having the frequency of the nerve impulses change and by having more or fewer nerves become involved.

It is worth emphasizing that this is it as far as classical ideas about sensory physiology go. There is far more unexplained about how the conscious mind forms a concept of the world from this than there are things we can explain with any assurance. Yet there is a confidence that the things we do know provide a basis for constructing a reasonably “good” picture of the external world.

It is necessary to refer the reader to other works for details (Rosen [15, 17, 18, 20] Mikulecky [22-25]), but the picture alluded to is called a *model* of the world if certain things are true about it. It is necessary to recognize that the mind has some system it uses to represent the external world. That system, which is called a *formal system*, comes into being as a result of sensory input of the kind just described. That is a very long, involved story. Let us recognize some relationship between things we observe changing in the external world that we *believe* to result from some cause, *causality*, the sensory data the mind receives from it, and some form of *encoding* of those signals into the formal system. This has to be true for we try to evaluate the effectiveness of all this by making inferences about changes

we experience in the external world by making *inferences*, that is to say, manipulations of the formal system by the mind, and then decoding the result of these inference in a way which allows some form of comparison with what was observed in the first place. There is a mathematical way of diagramming all this that involves mappings to represent the causal event we are attempting to explain, the encoding into the formal system, the inferential manipulation of the formal system, and the decoding to the external world. When the process being diagrammed works for us we say we have a model. In mathematical terms the diagram, called the modeling relation, commutes.

Now it is possible to deal with the myth of objectivity. The word myth is carefully chosen here because it is a belief that binds together so much in our way of looking at the world. It also is chosen to suggest that there are other ways of looking at the world. It is a myth because it ignores everything discussed here to describe how we form models. This is because only the formal system is subject to rigorous rules like those provided by logic. The formal system does not provide a way to accomplish the encoding, the decoding, the choice of formal system, nor the criteria for whether the modeling relation involving all these things really works as a description of the world or not.

1.2. Science and the myth of objectivity

Science has been our only real hope for an “objective” model of the real world. Unfortunately, it has had less than total success even though its success has been monumental. The reason lies in the discussion of human perception very quickly summarized here. The inbuilt need for the brain to supply so much to the raw sensory data is not capable of being overcome. The best we have been able to do is to work within a set of rather rigid rules and avoid those questions that required more freedom to explore. The result has been an explosion in the technological side of scientific thought and a withering away of any recognition of the value of keeping the philosophy up with the technology. As a result, the model science developed lost touch with the fact that it necessarily had to be encoding, using implication in a formal system, and then decoding to try to make models that worked. The criteria for what worked and what didn’t became more and more pragmatic until the success was the cause of an even larger failure. The scientific model works by suppressing the fact that there is an encoding and decoding from

the real world by making the formal system a substitute for the real world. This job has never been completed, but that does not weaken the *belief* that it will. The scientific model, even if unfinished, is widely, almost universally, accepted as a “largest” model; one which all other models derive from or fit into. It is only because of this that very brilliant minds could be seduced into accepting the myth of objectivity. Once the inescapable need for the encoding and decoding were forgotten, the necessary subjectivity built into the process could be ignored and finally denied.

Yet as the brain functions it clearly does not deal with raw sensory data. It processes and chooses according to what it has already learned and come to believe. The very act of trying to make objective measurements, reducing reality to numbers, abstracting severely, is the result of a very deeply entrenched belief structure. The belief structure has an inbuilt irony connected with it because it can not accept any evidence that would result in its having to be changed. Thus the quest for knowledge shuts out certain kinds of information and knowledge because it does not fit the model that has been so universally adopted. A good case can be made for supporting this. If we relax these criteria for what constitutes “scientific” information about “objective reality” there are all sorts of other belief systems that now have room to attempt to supply alternative models. The writings of skeptics about “quack” science and snake oil salesmen give all the evidence we need to know this is so. Thus we have are in a really difficult situation, there are risks to be taken or we stagnate. Notions like the idea that we have reached the “end of science” (Horgan [26]) are among the most pessimistic of these. The situation is not so grim (Mikulecky [22-25]). There are ways to proceed that do not throw the baby away with the bath water. One such approach is what will be outlined here. The approach being offered is not a replacement for the attempted largest model of classical science. That is taken as impossible from the start. Nor does it discard any of the useful achievements of classical science. What it does do is to knowingly step outside of those bounds and try to incorporate what was accomplished into a different framework that retains the knowledge that science is one of many belief structures and necessarily involves the encoding and decoding mapping from and to the real world like all other belief structures. Once this is done, there a case can be made that human minds are open to such a range of belief structures. The most obvious examples have to do with scientists who were or are also spiritual or even religious people. The previous statement assumes that religion is usually a more severe and

restrictive form of spirituality. The reason for assuming this should become clear later. Among the belief structures of importance are those that have their historical roots in cultures. This is a rich source of beliefs. It is most obvious in tradition directed cultures, but certainly not restricted to them.

1.3. Context dependence and self reference

Words generally do not have absolute meanings. The meaning depends on their context. This is the nature of semantics. It is the difference between semantics and syntax in language that serves as an analog model (a definition of what is meant by this will be forthcoming) for the problem we are dealing with. The formalists who sought for a truly objective way of understanding were seeking the analog of a language that had no ambiguity or context dependence. The idea that this argument is being written in English illustrates this issue very well. There is an old oral joke that asks how we spell the sound “fish”. The answer is ghoti. The reader can ascertain that it works. “gh” as in “tough” and so on. This type of ambiguity appears to a greater or lesser extent in all languages, but is replete in English. A language of pure syntax is not possible. The notion of language has built into it the diversity of relationships between syntax, the structure and algorithms that encode that structure, and syntax, the so-called meaning of that syntax. The latter aspect cannot be safely encoded into syntax. There is always more needed, an important part of which must be supplied by the human mind. Subjectivity really is going to be everywhere we look. It cannot be wished away. The crux of the problem has always been clear in language, but Escher did some nice things to bring it home in art. Escher’s art is worth a lot of discussion because it opens the Pandora’s Box of the things we learn about all this from visual sensation, but there is no space for that now. It has been discussed in this context at some length elsewhere (Hoffman [27]).

In language we have the whole set of paradoxes like the old example: “All Corinthians are liars. I am a Corinthian.” There are books full of related examples. There have been attempts to model these impredicatives, one of the latest being hyperset theory (Barwise and Moss [145]). It would be a distraction at this point to examine them further. Rather, the focus of this discussion will be an exercise carried out by Robert Rosen that began in the late 1950s and has been recently become a topic of discussion due

to Rosen's last two books (Rosen [19 ,21]). In these books the issue of self- reference and context dependence is central. A key part to the story is a paper written in 1972 (Rosen [16]). From discussion with readers of Rosen's books much is missed if the earlier paper is not also read. For that reason, the main development will be summarized here. This summary will provide the tools needed for dealing with these issues and supply an answer to the question "Why is the whole more than the sum of its parts". This answer will define another aspect of complex systems that exists side by side with their atoms and molecules. Many other definitions of complexity exist (Horgan [29], Mikulecky [22-25]).

2. An Introduction to Relational Systems Theory

Relational systems theory is a topic growing out of Robert Rosen's critique of classical systems theory as a part of the largest model science has created. (Rosen [18-21]). The mathematical form of the theory is a particular version of category theory that Rosen developed for this purpose. The inspiration for this radical new approach came from the Relational Biology created by Rosen's mentor, Nicholas Rashevsky [30], who has been called the father of mathematical biology. Relational Systems Theory combines a familiar form of systems representation and analysis, the use of block diagrams, with some concepts about causality taken from Aristotle.

2.1. Relational block diagrams

A simple representation of components to a system is the input/output block diagram. In this representation, each block represents an agent that effects a change on something, namely its *input*. The result of this interaction is some *output*. The abstract way of representing this is

$$f : A \rightarrow B \quad (2.1)$$

where f is the process that takes input A into output B. Clearly B can now become the input for some other process so that we can visualize a system as a *network* of these interactions. The relational system represents a very special kind of transition this way. Rather than break everything down in the usual reductionist manner, these transitions are selected for

an important distinguishing property, namely their expression of process rather than material things directly. This is best explained with an example. The system Rosen uses for an example is the *Metabolism-Repair* or [M, R] system. The process, f , in this case stands for the entire metabolism going on in an organism. This is, indeed, quite an abstraction. Clearly, the use of such a representation is meant to suppress the myriad of detail that would only serve to distract us from the more simple argument put this way. It does more because it allows processes we know are going on to be divorced from the requirement that they be fragmentable or reducible to material parts alone. In this way the existence of context dependence and self-reference is no longer a problem. That is what is gained at the expense of the reductionist focus on material parts. The idea is that if the whole is more than the sum of the parts there must be a meaningful way of representing this whole.

The transition, f , which is being called metabolism, is a mapping taking some set of metabolites, A , into some set of products, B . What are the members of A ? Really everything in the organism has to be included in A , and there has to be an implicit agreement that at least some of the members of A can enter the organism from its environment. What are the members of B ? Many, if not all, of the members of A since the transitions in the reduced system are all strung together in the many intricate patterns or networks that make up the organism's metabolism. It also must be true that some members of B leave the organism as products of metabolism. The usefulness of this abstract representation becomes clearer if the causal nature of the events is made clear. To do this it is necessary to consider the nature of *information* in a complex system. The usual notions of information derived for communications theory will not help. Something very different is needed.

2.2. Information as an interrogative. The answer to "why?"

Aristotle developed a set of causal answers to the question "why?" that opens an entirely new realm of description in systems. This set has four causes, material, efficient, formal, and final. They can be illustrated as answers to the question of causes for the existence of a structure such as a house. The answers to the question "Why is that house standing there?" are:

Material cause: The things that the house is made of, bricks mortar, wood, metal, glass, etc.

Efficient cause: That which assembled the materials into the finished house, the builders, manufacturers, etc.

Formal cause: The plans, blueprints that allowed the builder to assemble the materials into a particular form.

Final cause: The reason the house was built: To be a dwelling place.

The reductionist/mechanistic approach to systems has no place for this kind of information. It is considered irrelevant. The question answered is “How does something work. A house is actually uninteresting from this perspective since it is not a mechanism, but a mere “thing”. In the case of a living organism undergoing metabolism, both questions are interesting. The mechanisms of the organism constitute its *physiology*. Physiology combines anatomy and biochemistry with other information to answer how the organism does what it does. There is no interest in why these things happen because the question is not in the realm of classical science. What needs to be recognized is that the new information introduced by answering “why?” is of a very different kind and that the consideration of this information necessarily involves us in the knowing of what causes things to happen *independent* of the way they happen. Thus the two kinds of information will always be disjoint.

There is more. The relational statement that metabolism has the representation

$$f : A \rightarrow B \quad (2.2)$$

is a causal statement in harmony with interrogative information. It can be diagrammed

$$f : \rightarrow A \rightarrow B \quad (2.3)$$

where the broken arrow represents efficient cause while the solid arrow represents material cause. This gives f , metabolism, the interpretation: that which takes A to B .

2.3. Functional components and their central role in complex systems

In the context developed so far, the mapping, f , has a very special nature. It is a *functional component* of the system we are developing. A functional component has many interesting attributes. First of all, it exists independent of the material parts that make it possible. This idea has been so frequently misunderstood that it requires a careful discussion. Reductionism has taught us that every thing in a real system can be expressed as a collection of material parts. This is not so in the case of functional components. We only know about them because they *do something*. Looking at the parts involved does not lead us to knowing about them if they are not doing that something. Furthermore, they only exist in a given context. “Metabolism” as discussed here has no meaning in a machine. It also would have no meaning if we had all the chemical components of the organism in jars on a lab bench. Now we have a way of dealing with context dependence in a system theoretical manner. Not only are they only defined in their context, they also are constantly contributing to that context. This is as self-referential a situation as there is. What it means is that if the context, the particular system, is destroyed or even severely altered, the context defining the functional component will no longer exist and the functional component will also disappear.

2.4. The answer to “why is the whole more than the sum of its parts?”

It is the functional components of a complex system that provide and answer to that question. The semantic parallel with language is in the concept of functional component. Pull things apart as reductionism asks us to do and something essential about the system is lost. Philosophically this has revolutionary consequences. The acceptance of this idea means that one recognizes ontological status for something other than mere atoms and molecules. It says that material reality is only a *part* of that real world we are so anxious to understand. In addition to material reality there are functional components that are also essential to our understanding of any complex reality. This is Rosen’s most important breakthrough. It cannot be isolated from the other concepts used to formulate this argument. The context dependence, the self-reference, and the other ideas are all part of the conceptual framework. This conceptual framework *is not* modeled after the conceptual framework of reductionism. The two do not superimpose.

They stand side by side as ways of understanding. Thus there can be no largest model of reality. It takes at least these two ways of seeing the world to understand it.

2.5. **Reductionism and relational systems theory compared**

There is more to the difference between the two approaches than those points discussed so far. They can be summarized by comparison. The comparison can be presented as a difference between two types of system representation using the modeling relation already discussed. The system that is modeled using relational systems theory will be called *complex* and the system represented by the reductionist approach will be called *simple* (see Table 3.1). There is more to this distinction and the reader is referred to Rosen’s books for details.

The differences leave little room for ambiguity. Each has a meaning that involves all the others. The largest model of the reductionist approach disappears as soon as the encoding and decoding between the real world and our mind’s formal system are recognized. There is a more rigorous way of expressing this idea developed in detail in Rosen’s books. The whole is more than the sum of its parts because each real thing has its own semantics, its own context. This is a labile thing and can be lost or destroyed by taking the system down to its parts. There are shadows of this idea in reductionist thought that will be discussed in more detail. The entwining of the causal relations will become clear as the [M, R] system discussion is completed. Genericity is a concept Rosen develops in detail in his last book.

Table 3.1. The complex and the simple systems.

COMPLEX	SIMPLE
NO LARGEST MODEL WHOLE MORE THAN SUM OF PARTS CAUSAL RELATIONS RICH AND INTERTWINED GENERIC ANALYTIC ≠SYNTHETIC NON-FRAGMENTABLE NON-COMPUTABLE	LARGEST MODEL WHOLE IS SUM OF PARTS CAUSAL RELATIONS DISTINCT NON-GENERIC ANALYTIC = SYNTHETIC FRAGMENTABLE COMPUTABLE

The distinction between analytic and synthetic models is a very technical subject. It involves the mathematics of model representation in a space. The two kinds of models are different for any model that uses direct product spaces for one and direct sum spaces for others. This concept has been used to advantage in the discussion of quantum mechanical paradoxes among other things.

Fragmentability is the aspect of systems that can be reduced to their material parts leaving recognizable material entities as the result. A system is not fragmentable if reducing it to its parts destroys something essential about that system. Since the crux of understanding a complex system had to do with identifying context dependent functional components, they are by definition, not fragmentable.

2.6. The functional component is not computable

In order for the computer to do its work, it must be programmed with algorithms. Algorithms are the syntax of computation. There are no semantics. Yes, every time this comes up someone points to attempts to get computers to deal with semantics using algorithms. That is not the problem. The problem is in the context dependence and self-reference. These things are inherently not reducible to the algorithmic, syntactic form computers need in order to function. This has been an idea much discussed and to my satisfaction, has been laid to rest. Again I refer the reader to Rosen's books for a more involved technical exposition of the failures of the Church-Turing thesis. In a nutshell it claims that anything in the "real" world must be computable and this just is not so.

2.7. An example: the [M,R] system and the organism/machine distinction

The beginning of relational system theory was in the use of the [M, R] system to develop a model that made a firm distinction between the concept of organism and that of machine. The difference is manifest in the causal entwinement already identified as a characteristic of complex reality that is also missing in our mechanistic reductionist models.

The representation of metabolism as a functional component establishes the lack of any one to one correspondence between metabolism and any particular model made up of interconnected chemical reactions. The concept of metabolism cannot be reduced in that way. Surely, the interconnected

reactions, the diffusion processes, the other more intricate transport processes, and so much more are included in this simple representation. The idea behind relational models is to forsake the detailed physics and chemistry to recover what is lost when that detail is preserved. Hence, for the sake of a representation metabolism formally looks like $f: A \rightarrow B$.

The next step is to recognize the other functions that must also be part of an organism. These functions are necessary for the organism to do so many things that a machine cannot do. The organism adapts and learns and adjusts and even heals. No machine-like thing has these abilities. Machines can mimic them, but in a context that is incapable of uniting them all into a complex whole. The key feature that makes the organism different is that within the function called metabolism is a subset of process usually called catabolism and anabolism. Very simply these refer to synthetic and break down processes. Any chemist knows that reactions take place by one direction in the reversible process being greater in its effect than the other. When the two directions are proceeding at equal rates nothing happens and an equilibrium state has been reached. In the organism, things are much more involved. Synthesis occurs along a pathway involving one or more reactions and breakdown by an entirely different pathway. There are also network structures so that blockage in one place does not necessarily stop a crucial chemical change from occurring. This feature of the system would be a severe liability in spite of all the unique functions it provides if it were not for another self-referential aspect of the system. The functional component f itself has a source in the system. Figure 3.1 illustrates the causal relations that are entailed by metabolism itself and the second function,

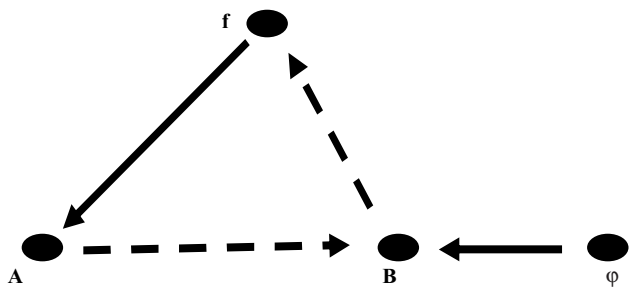


Figure 3.1. The primitive Metabolism-Repair system. Solid arrows depict efficient causation and dotted arrows depict material causation. The functional components f and φ effect the material changes from A to B and from B to f .

repair, symbolized by φ . The solid arrow from φ to B depicts the efficient causation that φ entails. The dotted arrow from B to f depicts the material causation resulting from the use of products of metabolism to *repair* f. Now there is both metabolism and repair, hence the name Metabolism-Repair System or [M, R] system.

The diagram is not very involved in its syntax, but its meaning, its semantics, are at the heart of what complexity theory has brought to the discussion. The transition from products of metabolism to the causal agent for metabolism embodies the answer to the question “Why is the whole more than the sum of its parts” if what we mean by parts is the usual idea of atoms and molecules. In this abstract model of the system, the functional component f has an entirely different character from the material entities symbolized by A and B. The latter are physical, material collections while the former is not capable of being reduced to the same kind of material things. It is a context dependent functional component that actually disappears when the system is reduced to its material parts.

Its mathematical character is that of a mapping between sets. The model treats this mapping and others we will see as if they had the same character as the sets. Relational models are a breakthrough in modeling for just this reason.

A note on the word “entailment” as it is used with these causal representations. We speak about entailment as the causation we are referring to in answer to questions of “why?” with regard to the existence of the entities in the system. In Figure 3.1 everything is entailed except the functional component φ . In this respect, what we have so far is not different from a machine. The machine lacks sufficient causal entailment to be self-contained. In order to have the system in Figure 3.1 something must entail φ from outside the system. This will always be the case with machines. The model of the organism, however, is not complete. There is, in fact, something in the organism that entails φ , its material cause is f and its efficient cause is β as is shown in the more elaborate model in Figure 3.2.

The idea that this may be the beginning of an infinite regression may be coming to mind at this point. If we were describing a machine with this model, such would be the case. In the organism it can be demonstrated in mathematical terms that given certain properties of the mappings we have introduced, the replication function β and the products of metabolism B are the same entity. The proof for this requires only that some of the mappings be 1:1. The one gene one enzyme relation is but one way of

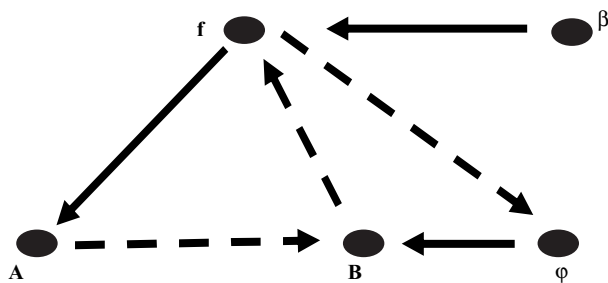


Figure 3.2. Repair is entailed by replication, another functional component.

fulfilling this mathematical requirement with real data from the organism as we know it. The proof then really identifies the mapping from β to f as having the same nature as that from B to f yet in one case it is an efficient causation and in the other case it is a material causation. The final model is shown in Figure 3.3. The need for an infinite regression is now gone. The organism differs from a machine in being *closed to efficient cause*. Notice that this is a necessary but not sufficient condition. The syntax of the diagram in Figure 3.3 is *not* a definition of organism. It could never be. Without the accompanying semantics it could represent many other things. In the context of those semantics, the diagram indeed represents the organism as a $[M, R]$ system that is closed to efficient cause. This

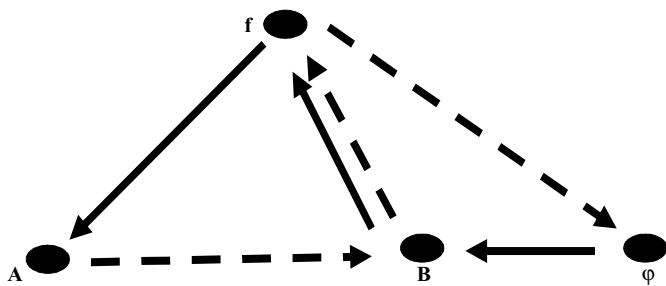


Figure 3.3. The relational diagram that is part of the way an organism is distinguished from a machine. The entailment is complete from within. It is closed to efficient cause. The syntax of the diagram is only part of the model. Some very sophisticated syntax *must* accompany it to make the model complete. Thus the requirement of the modeling relation to have semantics supplied to a purely formal system has been met. In the context of that semantics, the diagram represents an organism.

closure gives a formal model for Maturana and Varela's [31] concept of an autopoietic unity.

2.8. Relational models of mechanisms

Traditional science has produced a large number of models of mechanistic systems using the reliable techniques it has generated for this purpose. The use of relational models for mechanisms can be instructive but the utility of the traditional model makes this more of an exercise than anything else. Before relational modeling came on the scene, there were modeling methods for mechanistic systems that have an interesting and somewhat peculiar history. It is worthwhile to review some of that history to establish that the principles applied in the very abstract relational modeling of complex systems has certain parallels in traditional scientific modeling. This involves some ideas that never have become widely recognized. This, however, says nothing about their utility and their ability to reveal more about systems than the mainstream approach. The mainstream approach ultimately involves a largest model, namely a dynamics that comes from Newton's laws and the use of sophisticated differential and partial differential equations as equations of motion to obtain trajectories for any system, no matter how large or complicated. In recent years the ultimate has been achieved through the development of techniques to handle the most elusive of these trajectories that exhibit chaotic dynamics.

2.9. Newtonian dynamics is not unique; there are alternatives that yield equivalent results

An example of the Newtonian approach and its use of dynamics is the development of the trajectory of a particle. This involves quite a few assumptions as the actual particle's motion is abstractly encoded into an equation of motion. The mass of the particle is centered at point called its center of gravity. All that matters about the identity of that particle is that mass. Its motion can be deduced from its mass and location no matter what else we may know about it. Any particle with the same mass will move in the same way. This is so fundamental and so well accepted that these ideas about encoding into a formal model, stripped of everything else about the particle's identity, is taken as a description of reality by most scientists. To call it an abstraction may seem obvious to a non-scientist, but it is a description of reality in science. This is because of the utility it has provided.

This is a simple example, but the caveats apply far beyond the extent to which this example suggests.

There are three Newtonian Laws of motion and they can be stated simply here for the purpose of this discussion. First, The Law of Inertia simply establishes that particles remain at rest or in motion at a constant speed unless acted upon by some external force. The second, the relation between the mass of the particle, m , the external force, F , and the resultant change in the particle's motion, the acceleration of the particle, a is the equation of motion

$$F = ma \quad (2.4)$$

Much of physics is devoted to a host of very useful ways of using this equation. A simple one will serve to illustrate this use will be adequate.

The third law establishes the interaction between the force being applied and the particle by requiring that the exchange of force be reciprocal, the particle acts on the body supplying the force with an equal and opposite force.

The common force acting on all particles on our planet comes from the earth itself and is called gravity. Thus all particles fall towards the earth with the same acceleration (neglecting the friction of the air) as can be nearly perfectly established in a freshman physics lab. The acceleration due to gravity, g , is therefore a constant of the motion here on earth. The equation of motion is obtained by the use of two alternate ways of expressing the acceleration mass product

$$m \frac{d^2x}{dt^2} = mg \quad (2.5)$$

where x represents the position of the particle on a vertical line. This equation is easily solved by integration twice and by use of the initial conditions that the initial position is x_0 and the initial velocity is v_0 . The resulting trajectory of the particle is

$$x = x_0 + v_0t + \frac{1}{2}gt^2 \quad (2.6)$$

A different way of obtaining the same result comes from reasoning involving energy. This is a very simple example of the type of thermodynamic reasoning that will be developed in some detail shortly. By using the

exchange between the kinetic energy of the particle

$$KE = \frac{1}{2}mv^2 \quad (2.7)$$

and its positional or potential energy

$$PE = mgx \quad (2.8)$$

Looking at the initial state where

$$PE_o = mgx_o \quad (2.9)$$

and

$$KE_0 = 0 \quad (2.10)$$

And any other state, the Law of Energy Conservation requires that

$$PE - PE_0 = KE - KE_0 \quad (2.11)$$

Proper substitution and solving for x results in the same trajectory obtained by integrating the equations of motion.

The existence of an alternate to classical Newtonian dynamics has been known for a long time. It has not been given much significance. Even with the myriad of versions of so-called “complexity theory” (Horgan [27], Mikulecky [24,26]) there has been little discussion of the relevance of this lack of uniqueness of the seemingly largest model. The significance is more than trivial and deserves some discussion.

2.10. Topology, thermodynamics and relational modeling

Thermodynamics is usually a significant part of every physics textbook. It was Clifford Truesdell [32] who pointed out the formal difference between thermodynamics and all of the rest of physics. Thermodynamics is different for a very important reason that gets pointed out from time to time and then forgotten. The reason for the difference is that all of the rest of physics is the study of mechanism while thermodynamics is the study of system properties *that are independent of mechanism*. The reason that this is not as obvious as it might be is that thermodynamics has always been molded and shaped to be compatible with the study of mechanism. This attempt to force thermodynamics into the reductionist scheme has had some rather bizarre results.

Relational systems theory is also free of mechanisms and is based on a different kind of mathematics than is the traditional mechanistic approach. Mechanistic theories are centered on dynamics and make extensive use of calculus and differential equations. Much progress resulted in the past two or three decades due to the related progress in non-linear dynamics as formalism. This includes the entire field of “chaotics”.

Some important progress in the theory of chaotic systems is a result of Leon Chua’s [33] work on chaotic electrical networks. This is one piece of evidence of the utility of network theory in the broader study of systems.

Relational theories have a complimentary focus to that of dynamics. Rather than focusing on the details of the dynamics of the system’s parts, relations between parts are the center of attention. Often, these relations entail concepts that have little apparent connection with the dynamics of the parts. In a very real way, relational thinking is an extension of the thermodynamic reasoning described above. It says little or nothing about mechanism and particle motion. Instead it looks at the system’s *function*. In relational theories this can be formulated in terms of *functional components* formulated independently from the material parts of the system.

Network Thermodynamics is a transition between these extremes and has properties of both. It can give us an example of the application of dynamic systems theory to the making of models with the idea that represents a broader class of alternatives within the largest model. Other modeling methods such as cellular automata could have been chosen, but there are some features to the Network Thermodynamic formalism that allow certain points to be made during this discussion. The material aspects of the system are still the focus, but the functional aspects arise out of the *particular* organization of the *particular* system. In other words, the formalism has two distinct aspects, one constituting a general theory for formulating the thermodynamic aspects of complicated systems: (e.g., Meixner [34, 35], Oster, Perelson and Katchalsky [35, 13], Oster and Desoer [36], Oster and Perelson [39, 40], Oster and Auslander [38, 39], Perelson [42], Perelson and Oster [43], Peusner [4-12], Mikulecky and Sauer [46], Mikulecky [47]) the other a technique for modeling very complicated particular physical systems: (e.g., Blackwell [48], Breedveldt [49], Gebben [50], Karnopp and Rosenberg [51, 52], Koenig, Tokad, and Kesevan [53], MacFarlane [54], Rideout [55], Roe [56], and Thoma [57]).

It is in this latter aspect that its role as a facilitator for computer applications arises: (e.g., Wyatt [58], Wyatt, Mikulecky and DeSimone [59],

Mikulecky [47, 60-65], Mikulecky, Huf and Thomas [66], Minz, Thomas, and Mikulecky [67, 68], Peusner, Mikulecky, Caplan and Bunow [69], Oken, Thomas and Mikulecky [70], Seither, Trent, Mikulecky, Rape, and Goldman [71, 72], Seither, Hearne, Trent, Mikulecky, and Goldman [73], Talley, Ornato and Clarke [74], Thakker, Wood, and Mikulecky [75], Thakker and Mikulecky [76], Walz [77], Walz, Caplan, Scriven, and Mikulecky [78], White [79, 80], White and Mikulecky [81], Cable, Feher, and Briggs [82], Feher [83], Feher, Fullmer, and Wasserman [84], Fidelman and Mierson [85], Mierson and Fidelman [86], Fidelman and Mikulecky [87, 88], Cruziat and Thomas [89], Goldstein and Rypins [90], Horno, Gonzalez-Fernandez, Hayas and Gonzalez-Caballero [91, 92], Huf and Howell [93], Huf and Mikulecky [94, 95], May and Mikulecky [96, 97], Mikulecky and Thellier [98], Prideaux [99]). Both aspects of Network Thermodynamics bring in topology as a way of encoding the *organization* of the system along with its dynamics.

The application of Network Thermodynamics to chemistry has been mainly in the area of modeling chemical kinetics and the modeling of large chemical networks. This is because the problems that motivated its development were from biology. The modeling of chemical reaction networks is special because of the early onset of non-linearity in the equations. Recently, the usefulness of topological reasoning has become much more widely recognized in chemistry (Mikulecky [64]).

The second law of thermodynamics has been proved in a number of different ways, but the most elegant from a mathematical standpoint is the proof Caratheodry devised after Max Born lamented to him about the “roughness” of the Carnot Cycle proofs used before that. Caratheodry’s proof is a purely topological argument resting on one piece of experimental reality, namely the irreversibility of real processes (Mikulecky [47]).

Topology and mechanism free reasoning are paired in both Network Thermodynamics and in the kind of relational model devised to distinguish organism from mechanism. The idea that being closed to efficient cause is paramount is a completely mechanism free concept. The introduction of functional components allows us to even discuss process in a manner not dependent on the identification of the specific mechanisms entailed in those processes. If one considers how much scientific effort goes into trying to tie down illusive mechanisms, the impact of this should not be lost.

There is more to this comparison of the Newtonian Dynamics approach with other approaches. Modern non-linear dynamics started with topological reasoning as well. Poincare had the insight to realize that the requirement for complete solutions to nonlinear differential equations had to be sacrificed if any progress with real systems was to be made (Abraham and Shaw [100-104]). He introduced techniques that are now standard and have developed at a very significant pace in the area of chaotic dynamics. Phase plane diagrams, attractors, repellers, strange attractors, basins of attraction, etc. are part of a *qualitative* approach that is at the heart of this exciting field.

2.11. The mathematics of science or is all mathematics scientific?

The formal system into which real systems are encoded in the modeling relation is usually, as we have seen in this very brief discussion, some form of mathematics. Is all mathematics useful for this purpose? Once one understands that the candidate for largest model in the form of Newtonian dynamics has alternatives, the answer has to be in the affirmative. Yet most of the mathematical repertoire of the science curriculum has traditionally centered on the calculus and the solution of differential equations of motion to obtain trajectories. This is not the mathematics of science, but the mathematics of reductionist science. Even within reductionist science, topology and group theory have proven to be useful formal tools. Unfortunately, there has been much done using these and other mathematical tools that have not been universally understood or discussed because of the influence of reductionism on the curriculum used to prepare scientists mathematically. The usual practice of requiring calculus and differential equations clearly relates to the reason why these areas were developed in the first place. Newtonian dynamics and the calculus are all part of the same formal system and it is that formal system that obscured the fact that encoding was involved at all. The consequences of this are far from trivial. It is possible to become a highly respected and well-known contributor to the scientific literature without ever going outside the bounds of the traditional mathematical tools. The dominance of reductionism has more consequences than this, but in the context of this discussion the situation being described is as self-referential as it is disappointing to anyone trying to go beyond the

bounds of the reductionist paradigm. It is not really possible to describe systems holistically in that context.

There are many possible applications of mathematics outside the Newtonian dynamic framework that can be very helpful in making the problems that seemed to be outside the realm of science more accessible to a disciplined, rigorous approach. Even areas of reductionist science have been explored using other areas of mathematics for formalisms into which the real world can be encoded. The mathematical approach Rosen adapted for his study of the $[M, R]$ system is a version of category theory (Arbib and Manes [105]). Finally, topology and relational mathematics has been successfully used to reformulate Newtonian dynamics (Abraham and Marsden [106] and a significant portion of the formal aspects of chemistry (Oster and Perelson [40], Perelson and Oster [43]).

2.12. The parallels between vector calculus and topology

In their development of Network Thermodynamics, Oster, Perelson and Katchalsky [13] pointed out the usefulness of a number of additional mathematical tools in scientific modeling. Their history is interesting, but two earlier works mentioned stand out in particular. First the work of Kron demonstrates that all the partial differential equations of mathematical physics could be arrived at using network representations (a list of his works is referenced in their paper). The second is the work of Branin (1966) who carefully established homology between the vector calculus and topology.

3. The Structure of Network Thermodynamics as Formalism

Network Thermodynamics combines topology with analytical mathematics to model large complicated systems in a way which demonstrates the role of organization in dynamic models.

Network thermodynamics as developed By Oster, Perelson and Katchalsky [13] also used a representation for systems called bond graphs (Karnopp, and Rosenburg [51, 52]) which has become used by a number of scientists and engineers to deal with large complicated systems. Another

version of Network Thermodynamics developed by Peusner [4-12] uses the representation well known in electrical circuit theory and is amenable to computer simulation using circuit simulation programs such as SPICE as general purpose simulators (Mikulecky [47, 108]).

What Network Thermodynamics and physical systems theory have done is to demonstrate that the mathematical formalism that makes electrical circuit theory possible is applicable to all physical systems. The Newtonian approach can be replaced by energy conservation as shown in the simple example of a particle's trajectory. It is also possible to treat large complicated systems using network theory to obtain equations of motion and trajectories. The systems would be very cumbersome if treated as traditional boundary value problems even though that is what they are. Here is where topology plays the key role. Instead of setting up a large number of differential equations and struggling to make sure they are a system by matching a large number of boundary conditions, the combination of constitutive equations for the circuit elements and topology for keeping their connections straight is a real savings in effort. It also supplies a lot of mathematical structure that would never have been seen to apply in the boundary value approach.

3.1. Network thermodynamic modeling is analogous to modeling electric circuits

What makes the analog models possible is the recognition of two important ideas. The first is that in electrical circuit theory the circuit's topology or schematic is independent of what electrical circuit elements are used in the circuit. The two kinds of information can then be combined to arrive at a traditional systems description in terms of differential equations of motion and their resulting trajectories for the system. The second is that the circuit elements need not be electrical at all because of another key set of analogies. The pivotal topological unifier is in still another pair of mathematical relations the first of which has to do with conservation and the second with closure. These were introduced some time ago by Kirchhoff and are called Kirchhoff's laws. It is easy to show that even though these arise from conservation of charge and the closure of electrical potentials in closed loops, they apply to all physical systems because of the other conservation laws such as conservation of mass for mass flow and chemical

reactions and conservation of volume in watery solutions and other fluid systems.

This section will be devoted to summarizing the basic formal aspects of Network Thermodynamics. First its detailed character as a modeling tool will be developed then the more general theoretical aspects will be outlined.

3.2. The network thermodynamic model of a system

The network thermodynamic model of a system has two complimentary but distinct contributions. Their explicit formal independence and strict complementarity are one of the most striking aspects of the formalism. These two intertwined facets are the *constitutive laws for the network elements* and the *network topology*. The use of constitutive laws for the network elements is the way the physical character of each network element is represented abstractly. It is a common feature of the material world. The topology or connected pattern of these elements in a network is an independent reality about the system. The same topology can be realized for an infinite variety of collections of network elements. The same collection of network elements can be rearranged into a number of different topologies. The former fact is at the root of *Tellegen's Theorem* (Tellegen [109], Penfield, Spence and Duinker [41]), which is one of the major contributions of network theory to general systems theory.

3.3. Characterizing the networks using an abstraction of the network elements

The elements of any physical network can, with certain extensions for the non-linear systems, be classified from the relations between a simple set of observables, namely and *effort*, e , *across the element* and *flow*, f , *through the element* along with their time integrals called *momentum*, p , and *charge*, q (Oster, Perelson and Katchalsky [13,35]). These observables are associated with the networks elements in a manner that abstracts the effect of each of them as members of the system. The effort is some potential-like quantity's *difference across* the element while flow is movement of some entity *through* the element. The electrical manifestation of this general class of observables is the familiar voltage and current. These observables will be used to characterize the network elements uniquely. The analog

between these observables and the electrical network allows the entire body of electrical network theory to be applied more generally.

One thing that has to be clear is the fact that since thermodynamics is, by its very nature, true for all mechanisms it alone can do nothing to help us distinguish between realizable mechanisms. It does, on the other hand, serve us very well in distinguishing between realizable mechanisms and unrealizable mechanisms such as perpetual motion machines.

A second import point is that even though it is independent of mechanism, it can be used to derive mechanistic models when used in the proper context. A primitive example of this is the use of energy conservation (The First Law of Thermodynamics) to derive the particle trajectory mentioned earlier.

Thermodynamic reasoning is complimentary to mechanistic reasoning and there is an asymmetry in their relationship. We can describe a mechanism thermodynamically and thereby determine its realizability, but we cannot discern mechanism from thermodynamic descriptions alone (Callen [110], Hatsapoulos and Keenan [111], Prigogine and Defay [112], Tisza [113], Truesdell [114]). The early application of thermodynamic reasoning to non-equilibrium systems was in non-equilibrium thermodynamics (de-Groot and Mazur [115], Fitts [116], Prigogine [117]). This formalism was introduced into biology as a phenomenological approach to these complicated systems (Katchalsky and Curran [3], Caplan and Essig [118]).

3.4. The nature of the analog models that constitute network thermodynamics

The generality of Network Thermodynamics as modeling tool and theoretical formalism for all of physical systems theory is well established through the modeling relation. (Rosen [18,19]) The modeling relation can be used to define *analog* models. If the same Formal System is able to form commuting models for two or more Natural Systems, these systems are said to be analogs of each other and each could serve as a formal system for all the others. This is the case among physical systems with electrical networks being the representative Natural System. The fact that electrical networks were the first to be formalized extensively has made electrical network theory the source of models for a broader class of physical systems. To exemplify this analogy it is instructive to look at the constitutive relations in more detail.

3.5. The constitutive laws for all physical systems are analogous to the constitutive laws for electrical networks or can be constructed as the models for electronic elements

There are four possible binary relations among the network observables after the time integrals used to define charge and momentum are included (Oster, Perelson and Katchalsky [35], Peusner [12], Mikulecky [47]). Charge is the time integral of flow (or flow is a rate of change of charge) and the momentum is a time integral of effort. There are four distinct general network elements, each deriving their name from their electrical prototype.

- RESISTANCE relates effort to flow.
- CAPACITANCE relates charge to effort.
- INDUCTANCE relates momentum to flow.
- MEMRISTANCE relates charge to momentum.

Each of these network elements has its own unique interpretation with respect to how it handles energy. Foremost is the resistor, which is an idealization having the purpose to embody all the *dissipation* that goes on in a locality of the network element. Dissipation is the crux of irreversibility and the second law of thermodynamics. Systems in stationary states away from equilibrium are governed *totally* by resistance. Transient behavior comes from the time derivatives introduced by capacitance and inductance. In the Lagrangian formulation of networks, it is the resistance that represents the non-conservative aspects of the system (Mikulecky, Weigand and Shiner [119]). Capacitance is a form of energy storage without dissipation and is, therefore, also an idealization. The capacitor is a good model of a reversible isolated system in its behavior as it approaches its equilibrium potential. Since there is always dissipation in real processes, reality requires that there also be a resistor somewhere in series with the capacitor in order for the mathematics to be a faithful description of the system's trajectory. It is the capacitors that provide the dynamics in most models. This is a bit counterintuitive, since it is in equilibrium thermodynamics that these reversible (dissipation-less) energy transfers arise. Inductors are the idealized inertial elements and occur mainly in mechanics. The isomorphism between the differential equations for harmonic oscillators and LRC circuits has often been made. In the case of a weight bobbing up and down on a spring

fastened to a stationary object at its other end, the elastic spring is a capacitor, friction is the resistor, and the inertial force is the inductor. These seemingly simple analogical identifications link any physical system to the body of formal power resting in electrical network theory. This has been proven beyond a doubt to be a very significant formalism by its results, the myriad of achievements of modern electronics. This leaves the fourth element, the memristor. It also has analog physical realizations but they are rare (Chua [120]). So far, in all the applications encountered it has not been needed.

3.6. The resistance as a general systems element

Ohm's law is the binary relation between effort (voltage) and current (flow) in electrical networks. This *defining constitutive relation*, $e = ir$ defines the resistance as a simple proportion between effort and flow in linear elements. The effort, e , is actually a *difference* between the two electrical potentials *across* the element, $e = v_1 - v_2$, where v_1 and v_2 are the electrical potentials at each end of the element. The current, i , is the flow of charge *through* the element.

Fick's law does the same for diffusion or mass transport in physical systems. The flow through the element, j , is, as in the resistor, proportional to the potential difference, Δc , across the element, which in this case is a concentration difference. In this case the element may represent a membrane between two solutions or any other two regions separated by a diffusion barrier. The concentration difference, Δc , is the *analog* of the voltage and is the specific manifestation of the effort in this case. The mass flow, j , analogs the current and is the manifestation of flow in this specific case. Fick's law is

$$j = D \Delta c \quad (3.1)$$

where D is the *diffusion coefficient* of the flowing substance in the media making up the membrane or whatever space the diffusion traverses. As an analog to Ohm's law, Fick's Law can be rearranged to the form

$$\Delta c = j \left(\frac{1}{D} \right) \quad (3.2)$$

showing that $(1/D)$ is the analog of the resistance in the electrical case and is a specific manifestation of the generalized resistance. In other words, D

is a *conductance* and if Ohm's law is rearranged,

$$f = le \quad (3.3)$$

where the electrical conductance, $l = (1/r)$, is an analog of D in Fick's Law.

Poiseuille's Law describes bulk hydraulic flow of volume through a pipe,

$$Q = L_p \Delta p \quad (3.4)$$

where Q is the flow of volume through the pipe, p is the hydrostatic pressure and its difference across the pipe analogs the effort, and L_p , is the *hydraulic conductivity*. A trivial analog also can be made for chemical reactions, but they, in general, present a special problem which must be developed with more care.

3.7. The capacitance as a general systems element

Electrical capacitance, C , has both a static and a dynamic manifestation. In the static case it is the following relation between charge on the capacitor, q , and voltage across it, v :

$$C = \frac{q}{v} \quad (3.5)$$

To convert this to the dynamic form, the equation is rewritten as

$$q = vC \quad (3.6)$$

and then differentiated with respect to time:

$$I = C \frac{dv}{dt} \quad (3.7)$$

Hence, in networks where the dynamics are relevant, the capacitor relates flow to rate of change of effort.

The analogies developed among the different physical systems with respect to resistance are, in fact, true for all of network theory so that electrical network theory can be used as the prototype for all physical systems. To demonstrate this, consider, for example, some compartment of volume V . In it are n moles of some substance. Then, by definition, the concentration, c , of that substance in that compartment is

$$c = \frac{n}{V} \quad (3.8)$$

This can be rewritten as

$$n = Vc \quad (3.9)$$

which is analogous to the static definition of electrical capacitance if the amount, n , is analogous to charge, the concentration, c , is analogous to voltage, v , and the volume, V , is analogous to capacitance C . Taking derivatives with respect to time yields,

$$j = V \frac{dc}{dt} \quad (3.10)$$

which is completely analogous to the dynamic equation for electrical capacitance. This “osmotic” capacitance is applicable to any situation where a process changes the concentration in a compartment such as diffusion or chemical reaction.

A similar set of analogies hold for the pressure driven bulk flow either due to configuration of the system such as in the case of a U-tube or due to the compliance of the liquid. Either will result in some relationship between the system’s volume, V , and its hydrostatic pressure, p .

$$V = \gamma p \quad (3.11)$$

where γ is the analog of the electrical or the generalized capacitance. In dynamic form, after differentiation with respect to time,

$$Q = \gamma \frac{dp}{dt} \quad (3.12)$$

Similar analogies exist for the inductance and memristance, but since they have so far fewer occurrences in systems of interest, they will not be discussed here.

It is important to note that it is the capacitor that introduces a time derivative into a network’s mathematical description and thereby introduces *dynamics*. Purely resistive networks have a totally algebraic mathematical description and thereby describe stationary states away from equilibrium. Another way of saying this is that capacitors only are necessary during the transient phase of any simulation. When the system reaches a stationary state, they can be taken out of the model without consequence. The *voltage or effort source* may be seen as the limiting case of an capacitor with infinite (arbitrarily large) capacitance. In this case the effort approaches a constant value arbitrarily closely as its rate of change approaches zero.

3.8. The topology of a network

The collection of all the network elements cannot, by itself, constitute a network. There has to be something equally real to make it a functioning whole. The particular manner in which the elements are “wired together” must also be specified in some rigorous manner. This pattern of connections is the network’s topology. Each element has two ends, which was the basis for speaking of flow and effort as being observed through and across the element. Every element has to either be connected at its ends to another element or to “ground”. Ground is simply some reference potential, usually arbitrarily set to zero. These terms and concepts are obviously motivated by the structure of electrical networks and the analogy carries over to all physical systems. This is merely a way of saying that all physical systems can all be reticulated into a set of elements connected together. The basis for this statement has very deep roots in the structure of the underlying mathematics and has been though roughly established by Kron (Oster, Perelson and Katchalsky [19, 35] and Branin [107]).

3.9. The formal description of a network

In its most abstract form, a network consists of a set of nodes or vertices that are the connection points for the elements. This can be symbolized by the set of all vertices, V , such that any individual vertex, v_i , where $i = 1, 2, \dots, k$ and k is the total number of nodes, is a member of that set.

$$v_i \in V \quad (3.13)$$

The *network* is then a *relation* on the Cartesian Product, $N = V \times V = \{\text{all pairs } (v_r, v_s) \mid r, s = 1, 2, \dots, k\}$. In other words, any given network is a subset of N , the largest possible network with k nodes. This is an extremely abstract and formal approach to a very practical subject, but it is done to motivate the connection between network theory and the important relational mathematics generated by category theory (Rosen [17-19]). The relation that defines the network as a subset of N is the association of the pairs of nodes with the ends of elements in the network of interest. Once this association is made, each surviving pair of nodes defines a *link* or *edge* of the network. If each branch were to be represented by a line, the resulting structure would be a *linear graph* and if the order of the nodes in each pair were to be taken into account, the lines would have arrows from the first node to the second or the reverse. This latter case constitutes a

directed graph or *digraph*. The linear graph or the digraph corresponding to a network is an embodiment of that network's *topology* or *connectedness*.

For practical purposes, the constitutive relations describing the network elements and the topology of the network can be formalized independently and then combined to furnish a *solution* to the network. A network has a solution when all the observables in that network can be specified. The nature of the solution is a system trajectory. The formulation of the network is a set of coupled differential equations of motion.

Drawing the branches in the form of a connected set of lines or arrows with dots representing the nodes at the ends of the branches where the connections occur can diagram the formal representation of a network. This diagram is an application of graph theory, which is a part of topology.

By numbering the nodes and branches, another, equivalent representation is possible. This labeling system allows the construction of an *incidence matrix*, A . The incidence matrix has its columns numbered by the node numbers and its rows numbered by the branch numbers and the resulting matrix becomes an array of zeros and ones. In a linear (non-directed) graph only positive ones would appear and then only at the node/branch combinations where the node and branch were incident upon each other (the node is the end of that branch.) In a digraph a convention is adopted so that if the branch is incident on a node leaving the node it gets the opposite algebraic sign from the branch incident on a node entering that node. Using this definition, the linear graph representing the network's topology can be used to create the incidence matrix which is, in general a $b \times k$ array of zeros and plus or minus ones, where b is the number of branches and k the number of nodes. In turn, any incidence matrix has a unique realization in a linear graph. In other words they each can be used to generate the other.

The incidence matrix, by its nature, is a computational tool. This is easily demonstrated by the application of it as a representation of the networks topology to implement Kirchhoff's Laws [121]. These laws reflect two fundamental constraints on physical systems. They are consistent with the analogs developed among the different processes in that they apply equally well to all of them. The generalized version of the laws that had first been developed for electronic networks is:

Kirchhoff's Flow Law (KCL for Kirchhoff's Current law) states that at any node in the network all flows sum to zero given that incoming flows have the opposite sign from outgoing flows. This is a simple statement of conservation for the flowing quantity (mass, charge, volume in an

incompressible fluid, etc.). Using a vector of flows, $\mathbf{F} = (f_1, f_2, \dots, f_b)$ which lists the flow through each branch *in the identical order as they were listed in the incidence matrix*, KCL can be written in the form

$$\mathbf{A}\mathbf{F} = 0 \quad (3.14)$$

This matrix-vector product produces a list of the algebraic sums of flows at each node and nulls it. Clearly this can be easily programmed as a constraint in any program designed for solving these networks.

Kirchhoff's Effort Law (KVL for Kirchhoff's Voltage Law) states that around any closed loop in the network the algebraic sum of all efforts must be zero. This follows trivially from the fact that efforts are differences in potentials at the nodes across the branches so that there is one positive and one negative contribution from each node in the sum. It is equivalent to saying that if one made a hiking loop in the mountains, and stopped a number of times, the differences in elevations between the starting-stopping point and the rest stops will sum to zero unless there has been an earthquake. This also has a convenient, practical representation in terms of a vector of efforts, $\mathbf{E} = (e_1, e_2, \dots, e_b)$, a vector of node potentials $\mathbf{v} = (v_1, v_2, \dots, v_b)$ and the transpose of the incidence matrix $\mathbf{A}^*$,

$$(\mathbf{A}^*)\mathbf{v} = \mathbf{E} \quad (3.15)$$

This also is a useful representation of an important network property in a way that is readily utilized on the computer.

3.10. The formal solution of a linear resistive network

The real strength of network thermodynamics as a modeling tool is in its ability to simulate large, non-linear, difficult interacting physical systems. The following formal solution applies only to linear resistive networks and is here for mainly didactic purposes. Capacitors introduce network dynamics and lead to a set of state-space equations or equations of motion familiar to anyone who has studied linear dynamic systems. Non-linear systems require numerical work on a computer and are best handled by making the electrical analog and then simulating it on a strong circuit simulator such as SPICE. (Wyatt, Mikulecky and De Simone [59], Mikulecky [47], Mikulecky and Thomas [122], Tuinenga [123])

Using KCL,

$$\mathbf{A}\mathbf{F} = 0 \quad (3.16)$$

And the defining constitutive law for $\mathbf{F}$ is

$$\mathbf{F} = \mathbf{L}(\mathbf{E} - \mathbf{X}) + \mathbf{J} \quad (3.17)$$

where $\mathbf{L}$ is a diagonal matrix with the branch conductances on its diagonal, $\mathbf{E}$ is a vector of efforts across the resistors (conductances) in each branch, $\mathbf{X}$ is a vector of the force sources in each branch (in series with the conductors) and $\mathbf{J}$ is the vector of flow sources in parallel with each branch.

Using KCL,

$$\mathbf{A}\mathbf{F} = \mathbf{A}\mathbf{L}\mathbf{E} - \mathbf{A}\mathbf{L}\mathbf{X} + \mathbf{A}\mathbf{J} = 0 \quad (3.18)$$

or, using the relation between $\mathbf{v}$ and $\mathbf{E}$ above, rearranging,

$$(\mathbf{A}\mathbf{L}\mathbf{A}^*)\mathbf{v} = \mathbf{A}\mathbf{L}\mathbf{x} - \mathbf{A}\mathbf{J} \quad (3.19)$$

Defining an *admittance matrix*, $\Lambda = (\mathbf{A}\mathbf{L}\mathbf{A}^*)$, which has an inverse, Λ^{-1} , yields the solution to the network in terms of the vector of node potentials, $\mathbf{v}$

$$\mathbf{v} = \Lambda^{-1}\mathbf{A}\mathbf{L}\mathbf{x} - \Lambda^{-1}\mathbf{A}\mathbf{J} \quad (3.20)$$

When time derivatives are introduced by capacitances (or, more rarely, inductances which are inertial in character in mechanical systems) this becomes a version of the well-known *state vector* equations or equations of motion which are the substance of dynamics (DeRusso, Roy, and Close [124]). What is *special* about formulating the equations of motion in this manner is that the relationship between the constitutive relations describing the network elements and the topology of the network are explicitly identifiable via their mathematical encodings in the diagonal conductance matrix (and diagonal capacitance and inductance matrices in the more general case) and the incidence matrix respectively. The formal solution for a non-linear network involves replacing the linear constitutive laws by non-linear versions, so that the generalization of Ohm's law relating effort to flow takes the form(s)

$$\mathbf{E} = \mathbf{R}(\mathbf{F}) \quad (3.21)$$

and

$$\mathbf{F} = \mathbf{L}(\mathbf{E}) \quad (3.22)$$

where the conductance function, $\mathbf{L}$, and the resistance function, $\mathbf{R}$, are now non-linear functions of the flow and effort respectively (Chua [125], Chua and Lin [126]). These non-linear constitutive relations still occur for each

network element separately, but nevertheless they clearly complicate the mathematics significantly. The extreme case is the simple non-linear circuit containing one non-linear element (resistor) among a group of standard linear elements (resistors, capacitors and an inductor) that has been studied in detail because of its chaotic behavior it has been described using the *double scroll attractor* (Chua and Madan [32]). The study of large non-linear networks has been furthered most effectively in the electronics field and it should be no surprise that some of the most useful tools for dealing with these networks were developed in that field.

3.11. The use of multiports for coupled processes: the entry to biological applications

The multiport or n-port is a device which models coupled flows (Chua and Lam [127], Mikulecky [128]). Once again, for didactic purposes, the simpler linear case will be demonstrated. The more useful (and complicated) non-linear cases can easily be simulated on the computer, using SPICE as described later in this review.

3.12. Linear multiports are based on non-equilibrium thermodynamics

The linear 2-port is simply a model of the phenomenological equations of linear non-equilibrium thermodynamics. Its general structure is shown in Figure 3.4. If we replace symbol J for the flows with the symbol F and likewise identify the thermodynamic forces (X s) with efforts (E s) in the



Figure 3.4. A linear 2-port network element.

conductance format it has its mathematical representation as

$$\begin{aligned} \mathbf{F}_1 &= \mathbf{L}_{11}\mathbf{E}_1 + \mathbf{L}_{12}\mathbf{E}_2 \\ \mathbf{F}_2 &= \mathbf{L}_{21}\mathbf{E}_1 + \mathbf{L}_{22}\mathbf{E}_2 \end{aligned} \quad (3.23)$$

or in the resistance format as

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{R}_{11}\mathbf{F}_1 + \mathbf{R}_{12}\mathbf{F}_2 \\ \mathbf{E}_2 &= \mathbf{R}_{21}\mathbf{F}_1 + \mathbf{R}_{22}\mathbf{E}_2 \end{aligned} \quad (3.24)$$

This 2-port is easily generalized to an n-port device. Figure 3.4 shows a visualization of the 2-port as a network element. By rearranging the equations as follows

$$\begin{aligned} \mathbf{E}_1 &= (\mathbf{R}_{11} - \mathbf{R}_{12}) + \mathbf{R}_{12}(\mathbf{F}_1 + \mathbf{F}_2) \\ \mathbf{E}_2 &= \mathbf{R}_{21}\mathbf{F}_1 + (\mathbf{R}_{22} - \mathbf{R}_{21})\mathbf{F}_2 \end{aligned} \quad (3.25)$$

The network can be redrawn with the coupled flows “injected” into the resistors $\mathbf{R}_{12}$ and $\mathbf{R}_{21}$. This produces a purely resistive network, but it is disjoint. If we recognize the Onsager reciprocal relation [128,129],

$$\mathbf{R}_{12} = \mathbf{R}_{21} \quad (3.26)$$

The network becomes connected as shown in Figure 3.5. There is a unique relation between the Onsager reciprocity condition and the topological connectedness of the network representation!

This ability to reticulate the linear 2-port into simple resistors does not exist for non-linear 2-ports. This is the heart of why the reductionist approach

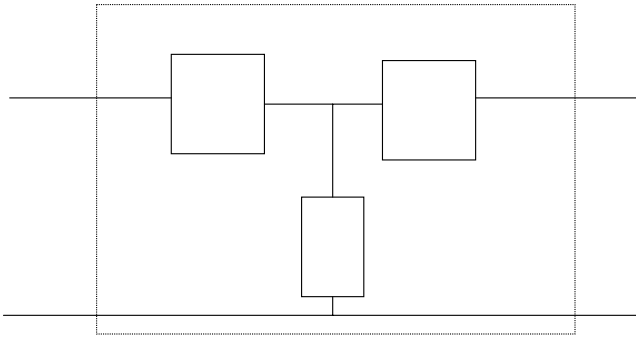


Figure 3.5. The linear resistive 2 port is a simple resistive circuit.

was so healthy while science had to rely mainly on linear models. The reduction ceases to be possible in non-linear systems. This is of importance in another, fundamental way in the network formulation of equilibrium thermodynamics as well.

For Biology, the n-port is the thermodynamic answer to the energetics of the biomass on the planet. It is the only way we have to model the marvelous processes that create structure and organization while the second law of thermodynamics collects its price – the continual production of entropy. In n-port coupled systems, any number of ports may be involved with the creation of negative entropy as long as the *overall net* entropy production is positive.

Because of the nature of physical systems, it is only the resistor elements that require multiport representation. Capacitances are modeled in the same manner when these devices are used. Linear multiport networks generate the same type of linear equations of motion as in the simpler case and present little further complication. It is easy to show that the algebraic structure of a network's solution remains invariant as the network is changed from a simple 1-port network to an n-port network. Furthermore, the steady state involves only resistive n-ports and sources.

A specific example of a 2-port network element would be the reaction-diffusion 2-port. In this case, substances A and B are diffusing across a region that also catalyzes their chemical interconversion. Diffusion is analogized by simple resistors while the first order reaction kinetics are analogized by controlled sources in the form of *unistors* (Mason and Zimmerman [132]). Unistors are special network elements that rectify flow completely and have a constitutive law that is of the form

$$\mathbf{F} = \mathbf{k}v_1 \quad (3.27)$$

where $\mathbf{k}$, and v_1 is the potential at one end of the unistor.

Another example is the *convection – diffusion* 2-port that is used to represent an aspect of biological membrane transport. Its constitutive equations are (Kedem and Katchalsky [133-135])

$$\mathbf{Q} = \mathbf{L}_p(\Delta\mathbf{P} - \sigma \Delta\Pi) \quad (3.28a)$$

$$\mathbf{F} = \omega RT\Delta\mathbf{c} + \langle \mathbf{c} \rangle (1 - \sigma)\mathbf{Q} \quad (3.28b)$$

where $\mathbf{Q}$ is the bulk flow, $\mathbf{F}$ is the diffusion flow, $\Delta\mathbf{P}$ is the hydrostatic pressure difference across the membrane, $\Delta\Pi$ is the osmotic pressure

difference across the membrane, L_p is the hydraulic conductivity, ωRT is the permeability of the membrane to the diffusing substance, Δc is the concentration difference of the diffusing substance across the membrane, $\langle c \rangle$ is the average of the concentrations in the baths bathing the membrane, and σ is the reflection coefficient for the diffusing substance in this membrane

4. Simulation of Non-Linear Networks on Spice

The mathematical difficulties encountered when the network elements are non-linear are most easily handled by computer simulation on SPICE. The simple examples given above should not be misinterpreted. They are demonstrations of a method that is more useful as the problem gets more difficult. The ultimate difficulty is not in the size of the network but in the presence of non-linearity in the constitutive relations of one or more elements. As in any other method for dealing with such difficulties, the computer and numerical analysis come to the fore. In the case of network thermodynamic models, there is a distinct advantage.

As the explosion in chip technology became a central theme in electronics, a method for the design and simulation of these elaborate, large non-linear networks had to be developed. The development of simulators is a saga worthy of review in its own right. For our purposes it is sufficient to relate that one outcome of this enormous effort has become well entrenched and almost a standard as such programs evolved. The circuit simulator *SPICE* (Simulation Program with Integrated Circuit Emphasis) developed in the EE and Computer Sciences department at UC Berkeley is now used extensively in the industry (Chua and Lin [126], Tuinenga [123]). Its use as a general physical systems simulator was developed by Thomas and Mikulecky [129] after they did some initial simulation work on the French program AZTEC. Over the years, a myriad of biochemical, physiological, and pharmacological systems were successfully simulated using this program.

The linear elements are analogized as resistors and capacitors as described above. The non-linear elements are represented by devices called *controlled sources* that allow the simulation of non-linear resistors and capacitors as well as the more complicated chemical reactions that are often very non-linear. In biochemistry, special kinds of non-linearity arise in

enzyme-catalyzed reactions. This involves the Michaelis-Menten class of reaction mechanisms and the various forms of inhibition interactions they entail.

4.1. Simulation of chemical reaction networks

The simulation of chemical reaction networks on SPICE has had significant applications. The methodology is rather simple (Wyatt [58], Wyatt, Mikulecky and De Simone [59]). The most extensive of these applications is in the area of biochemical/pharmacological networks (Thakker, Wood, and Mikulecky [75], Thakker and Mikulecky [76], Walz, Caplan, Scriven, and Mikulecky [78]). Let's look at an example from biology. This particular system, folate metabolism, is an important one in the synthesis of nucleic acids on the way to making building blocks for DNA and RNA. For that reason it plays an important role in cancer chemotherapy. (Seither, Trent, Mikulecky, Rape, and Goldman [71, 72], Seither, Hearne, Trent, Mikulecky, and Goldman [73], White [79, 80], White and Mikulecky [81]).

The particular biological flavor of this problem manifests itself in many ways, not the least of which is the nonlinearity of the kinetics for each reaction step and the many interactions between constituent chemical entities in the form of various types of inhibition (competitive, non-competitive, etc.)

Included in some of these studies is the parallel capacity to simulate the distribution and transfer of materials in compartmental systems, a specific application of reaction diffusion systems theory and/or batch processing theory.

4.2. Simulation of mass transport in compartmental systems and bulk flow

As demonstrated in the works cited in the previous section, simulations combining non-linear reaction networks with mass transport introduce no additional complication. The use of *multiports* enables the simultaneous simulation of different, interacting processes.

The use of multiports allows chemical reaction and mass transport to be coupled with bulk flow. In chemical operations this may be a very helpful feature.

4.3. Network thermodynamics contributions to theory: some fundamentals

As late as 1973 Callen [136] wrote:

“In short my thesis is that thermodynamics is the study of those properties of macroscopic matter which follow from the symmetry properties of physical laws, mediated through the statistics of large systems.

Two considerations contribute at least to the *a-priori* plausibility of this construction. Firstly, it rationalizes the peculiar nonmetrical quality of thermodynamics.”

Since that time there have been successful attempts to rectify the situation. Among them is Network Thermodynamics and Thermostatistics that have made this rationalization totally unnecessary. In particular, using this approach, the assumption Callen makes that these things must come about through statistics is shown to be totally uncalled for.

4.4. The canonical representation of linear non-equilibrium systems, the metric structure of thermodynamics, and the energetic analysis of coupled systems

We now know that network thermodynamics is the canonical representation of linear non-equilibrium thermodynamic systems. The Onsager/Prigogine representation is a reductionist partial description. When the more holistic approach of Network Thermodynamics extends the non-equilibrium thermodynamic formalism, all steady state systems are complete circuits with resistors AND sources. The formalism Onsager and Prigogine developed only studies the resistors. That practice caused Tellegen’s Theorem and the accompanying mathematical structure of the state space to be missed. This is profound! Onsager’s formulation of non-equilibrium thermodynamics was found lacking by Callen, Tisza and others due to its affine coordinates. Peusner showed that *every* Onsager system has a unique embedding in an orthogonal, higher dimensional system and that those coordinates were precisely those dictated by the network representation! In other words, the simple 2-port networks representing the linear two-force/two-flow systems are more than just a convenient representation. The three resistors in this “T” network identify a set of orthogonal coordinates into which every affine Onsager system can be imbedded. This provides

a *common metric* for measuring distance in entropy/energy spaces as far from equilibrium as we like (Mikulecky [47]).

The geometry and the energetics are tightly coupled here. Kedem and Caplan [137] showed that there was a transformation of variables in any Onsager system (linear phenomenological description) that yields an important geometric invariant, q , the *degree of coupling* (Caplan and Essig [118]). For any linear 2-port

$$q = \frac{L_{12}}{(L_{11}L_{22})^{1/2}} = -\frac{R_{12}}{(R_{11}R_{22})^{1/2}} \quad (4.1)$$

The sign of q depends on whether the coupled process are effectively moving in the same (positive) or opposing (negative) directions. Using the constraint of positive definiteness on the coefficient matrix (the $\mathbf{L}$ matrix) imposed by the second law of thermodynamics, the value of q is also restricted

$$-1 \leq q \leq 1.$$

The maximum efficiency of any linear energy conversion device is a function of q alone:

$$\text{Maximum Efficiency} = \frac{q^2}{(1 - (1 - q^2)^{1/2})^2} \quad (4.2)$$

In the embedding described above, q is the parameter that determines the angle of “tilt” in order for the affine Onsager coordinates to fall onto the orthogonal set of planes making up the canonical orthogonal coordinate system. Thus q is indeed an important invariant of the system in this way also.

4.5. Tellegen’s theorem and the onsager reciprocal relations (ORR)

Lars Onsager [130, 131] won the Nobel Prize a while back in part for his work on non-equilibrium thermodynamics and in particular, his proof of the “reciprocal relations”. His proof used statistical physics in the domain of *fluctuations* around equilibrium. The decay of those fluctuations was described by the linear phenomenological laws and the use of statistical physics was thought necessary to obtain the proof. Peusner [11, 12] was able to accomplish the proof much more simply, accurately and directly

using Tellegen's Theorem which is a result at the most basic level in network thermodynamics. He systematically matched elements of his proof with that of Onsager to show that all the molecular statistics could have been ignored since the necessary and sufficient elements of the proof are topological properties that Onsager had implicitly assumed and which were in no way dependent on molecular statistics.

Tellegen's theorem is, in its simplest form, a statement of power conservation in a complete network. To be complete, the network must have energizing sources or charged capacitors to make it work. The simplest possible example is the series circuit containing a single resistor and a constant source of current or voltage. Tellegen's Theorem is then

$$J_s X_s + J_r X_r = 0 \quad (4.3)$$

For any network the vector of flows and the vector of efforts (forces) are orthogonal:

$$\mathbf{J}\mathbf{X} = 0 \quad (4.4)$$

In other words, the power dissipated by the resistor and the power supplied by the source are equal and by using the appropriate sign convention of opposite sign so they sum to zero. This looks trivial, but it has a very general form for any network that can be derived using Kirchhoff's laws and the network topology, *independent of the identity of the circuit elements*. For that reason, the *quasi power theorem* is easily proved as well.

In this version there are two different networks (starred and unstarred) with exactly the same topology. If we take a vector of flows from one and a vector of forces (efforts) from the other, these vectors will always be orthogonal!

$$\mathbf{J}\mathbf{X}^* = \mathbf{J}^*\mathbf{X} = 0 \quad (4.5)$$

These can either be two different networks or the same network at different times. Using this Onsager's reciprocity theorem can be proven without the use of statistical thermodynamics or near equilibrium assumptions other than that the system is still in the linear domain.

It had been known by experimentalists for about a century, that the reciprocal relations (in particular Saxen's relations (Miller [138, 139]) held in non-equilibrium situations far enough from the domain of Onsager's proof (fluctuations around equilibrium) to make his proof seem odd, at

best. Peusner's topological proof not only did not need or use statistical physics, but also was valid *everywhere* in the linear domain far beyond the domain of Onsager's attempt. Thus it was in complete harmony with the myriad of experimental results.

Furthermore there is an intimate link between the ORR and the ability to have a simple, topologically connected network element representing any linear non-equilibrium (Onsager) system. In fact this is the canonical representation that yields the metric in entropy or energy space. This is but one more demonstration that topological (non-mechanistic) considerations underlie many fundamental relations in these systems.

5. Relational Networks and Beyond

5.1. A message from network theory

Network Thermodynamics is presented here as a transition between the classical mechanistic approach to systems theory and a more modern relational approach. It is used as a very special example of the class of modeling techniques that can be incorporated into the largest model of dynamic systems theory. Another approach might be cellular automata, for example. The Network Thermodynamic approach illustrates the way organization in the form of network topology plays a key role in the formalism that leads to a set of equations of motion and resulting trajectories. Network thermodynamics, as it has been applied in biology, has been intended to focus on the complexity of these systems. Yet it uses standard mechanistic observables and results from a reductionist approach to synthesize elaborate models of these complicated interacting systems. Complexity theory means many things to many people (Horgan [29], Mikulecky [24, 26]), but one common thread is that it is a reaction to the effect of *reductionism* on modern scientific thinking and practice. Complexity theory uniformly deals with a statement that has become one of its central dogmas:

“The whole is more than the sum of its parts”

To use the approach to complexity put forth by Robert Rosen [18-21] (see also Mikulecky [24, 26]), the question “Why is the whole greater than the sum of its parts?” needs to be answered. Its answer is deep and

certainly not obvious. It has had partial answers from the growing number of scientists showing an interest in complexity research. Those partial answers all involve the idea that “emergent” properties arise in complex systems and that these properties somehow arise out of the system’s material parts, but do not map to them in any clear way. Had these emergent properties been easily predicted from the parts of the whole, the notion of emergence would never have arisen. Instead, these emergent properties often either involve surprise or a suspicion of error or both. It is possible to illustrate a simple version of this with multi-port networks. One clear example is in network thermodynamic models of salt and water transport through epithelial membranes such as those that line the gut, kidney tubules, and gall bladder.

5.2. An “emergent” property of the 2-port current divider

Emergence is a central concept in complexity theory. There are at least two distinct kinds of emergence that can be identified and their difference is important. There is emergence in the real world whenever evolution or developmental processes occur. In the modeling of the real world, emergence is a phenomenon of the modeling process itself when unexpected results come from a model. The case considered here will deal with emergence as a model is changed from a simple set of 1-ports to 2-ports.

The model was created to show how sodium transporting epithelial membranes can transport large volumes of isotonic fluid across the wall of structures such as the gall bladder, the small intestines, and kidney tubules. This model involves a series/parallel combination of convection-diffusion 2-ports and a source representing the active transport pump across the basolateral cell membranes. The tissue as a whole is capable of transporting isotonic sodium chloride from the lumen to the blood across the tissue even if no gradients exist. The new, *emergent*, property involved is *the 2-port current divider principle*. In the case of simple 1-ports, a current source feeds into the node connecting a pair of resistors. The current splits up according to the resistances relative values. In the 2-port case, if the input is solute flow and the two 2-port elements are not identical, volume will flow across the system from zero potential to zero potential at the other end. This new property of the 2-port system is indeed an emergent property and tells

us that the world of physical multi-port devices is as capable of producing interesting new phenomena, as was the case when the multi-port was introduced in electronics. Some variations on this theme involve a biological fuel cell and other interesting devices (Mikulecky [47]).

This emergent property was shown to both reinterpret established results and to predict new findings in the epithelial tissues that had been the motivation for the analysis (Fidelman and Mikulecky [88], Mikulecky [47, 108, 122, 128]) of a network model for coupled solute and volume flow through an epithelium.

What these theoretical developments show, in a consistent way, is that even in physical systems that are derivable from the Newtonian Paradigm's "largest model" there is much of the overall system's properties that is lost unless a more holistic approach is used. Network Thermodynamics, by including the energizing sources along with the dissipators uses the entire system in its considerations. By so doing, the formalism reveals some of the most interesting properties a system may possess. In *every* case these properties arise *independent of* the basic physics in the constitutive relations. These properties are primitive *relational* properties and provide a basis for understanding the more abstract relational approaches in complex system theory.

This brief sketch of the network thermodynamic formalism should make clear the reason why scientists who believe that the study of the world can only be done by certain "safe" methods, namely the repertoire of methods included in classical dynamic systems theory. The very mathematical training of most scientists speaks to the idea directly. Newton's calculus and its application to systems of differential equations has become the mainstay of scientific mathematics. Topology, Category Theory, and related "relational" mathematics are not familiar tools. This causes a mind set and an unconscious belief in the scientific method as we know it that is strong enough to make us feel comfortable that the answers we seek will come from learning to use this mathematics and this method as well as we can. This limits the universe of discourse. Complexity science emerged because this universe of discourse was too small. Too many interesting questions fall outside of it.

Network Thermodynamics took a new approach to systems theory and demonstrated that we were missing some really important ideas. It is now time to incorporate Network Thermodynamics into systems theory as a way

of dealing with mechanism within the broader scope of relational systems theory.

5.3. The use of relational systems theory in chemistry and biology: past, present, and future

When Kedem and Katchalsky [140, 141] introduced non-equilibrium thermodynamics into biology in the late 1950's they were among a handful of others who saw the need for this formalism in the study of living systems. Prior to this, equilibrium thermodynamics had some usefulness in trying to understand the energetics of living systems, but since their homeostatic nature made them much more like stationary states away from equilibrium, the theories for equilibrium were not very helpful.

Linear non-equilibrium thermodynamics itself had more of an impact in the realm of rethinking the conceptual framework for asking and answering energetic questions about living systems. From the very beginning of this rethinking it was clear to people like Katchalsky and Prigogine that the real breakthroughs would come when the formalism could be extended to nonlinear systems. This may be part of the reason why the conceptual insights gained using linear non-equilibrium thermodynamics never became as widely understood as they needed to be. As a result much time and energy goes into our modern discussions of complex systems in order to fill in gaps in the overall picture.

What were some of the conceptual changes that resulted from the non-equilibrium formalism? First and foremost, the entire set of rich conclusions about the nature of coupled systems and the appropriateness of this model for understanding how life was a direct result of the requirements imposed by the second law of thermodynamics. The highly interactive nature of living systems arises because of circumstances that require a response to a steady input of solar energy. The way the system responded to this energy throughput was by organizing better and better ways to keep this energy from becoming accumulated heat. Through coupled processes, the heat flow was channeled through biomass and a cooling effect was achieved which was part of the stabilization of the atmosphere. That atmosphere in turn provided a milieu that could sustain life.

A second realm of advances came for more technical reasons. Kedem and Katchalsky chose the realm of membrane transport for their first round

of applications and, in particular, began by showing that without modeling the osmotic transient with a set of coupled equations, there was no way to fit the experimental curve describing the swelling and shrinking back to normal volume of a cell exposed to a slightly hypotonic solution due to *permeable* solutes. The role of the coupling coefficient, the reflection coefficient, in explaining some of the more difficult osmotic effects in tissue then followed from that.

Today, this information has become an integral part of every physiology text that deals with osmosis in a careful way.

A very useful result was in the application of *Curie's Principle* to the "relational" explanation for active transport. The simplest representation of active transport using the linear non-equilibrium formalism is as follows:

The linear domain of the system can be modeled phenomenologically by the equations,

$$J_r = L_{rr}A + \mathbf{L}_{rs} \cdot \Delta\mu_s \quad (5.1)$$

$$\mathbf{J}_s = \mathbf{L}_{sr}A + L_{ss}\Delta\mu_s \quad (5.2)$$

where J_r is the flow (rate) of the ATPase reaction, A is the reaction affinity, $\Delta\mu_s$ is the chemical potential difference across the membrane for the solute being actively transported, $\mathbf{J}_s$ is the solute flow, and the L 's are the coupling coefficients. The bold symbols are vectors having both magnitude and direction, while the others are scalars. The chemical reaction flow and force are scalars while the mass transport flow of solute through the membrane and its thermodynamic driving force, the chemical potential difference across the membrane, are vectors. In order for the equations to make mathematical and physical sense, the coupling coefficients must be vectors as well. The dot in $\mathbf{L}_{rs} \cdot \Delta\mu_s$ is the vector dot product that results in a scalar. These "vectorial" coupling coefficients represent something about the structure of the space in which all this is happening. Either that space is asymmetrical so the coupling coefficient may be non-zero or the space is symmetrical and there is no coupling between mass flow and reaction, or, in other words, no active transport.

Curie's original statement could be translated and paraphrased as "Without the breaking of symmetry, nothing happens." It has a rigorous physical manifestation in non-equilibrium thermodynamics in that flows and forces of different tensorial character do not couple in an isotropic space (deGroot

and Mazur [115], p. 31-33). In this case we can apply it by stating that the necessary and sufficient condition for active transport is that a chemical reaction takes place in an asymmetric space. The link between the two versions of Curie's principle was solidified by DeSimone and Caplan [142, 143] although its meaning was understood by Kedem and Katchalsky in the late 1950's. To this day it has had little impact on biologists. Reductionism has dictated a need to see a molecular mechanism to "explain" the phenomenon before they believe that they understand it.

Let us consider a relatively simple man-made experimental system to see what one mechanistic realization of this general principle can look like. Consider two cases of the system consisting of an enzyme imbedded in a membrane separating two solutions. In case a the enzyme is evenly distributed throughout the membrane while in case b it is only in one half. If the enzyme catalyzes the reaction $A \rightarrow B$ and each of the two solutions contains equal concentrations of A and B such that the ratio of concentrations is not the equilibrium ratio, the membrane bound enzyme will catalyze the reaction as A diffuses to enzyme sites in the membrane and B diffuses away. In the symmetric case, the diffusion paths are equal from either bathing solution so that the concentrations of A and B on both sides remain equal to each other even as the concentration of A diminishes and that of B is increased.

In the asymmetric case, the diffusion path through the membrane is much greater on one side than it is on the other. In that case the conversion of A to B one side will lag that on the other side and a gradient of A in one direction and B in the other will be created. Hence a gradient can be created *iff the reaction is in an asymmetric space*. (The same effect could be achieved by distributing the enzyme evenly throughout the membrane and shrinking the pores on one side). A close examination of every known scheme for active transport, either in theoretical models or in experiments obeys this symmetry/asymmetry principle.

This example shows more than an application of the non-equilibrium thermodynamic formalism. It shows the alternative way of approaching modeling in systems. It demonstrates that this type model can lead to principles which determine the conditions for important processes like active transport to exist. Yet, due to its non-mechanistic, phenomenological character, it does not get the status of an "explanation" among biologists. The new relational biology (Rosen [15-19]) develops this line of thinking further and points the way to a way of understanding the complexity of living

systems as distinct from the simple mechanisms that have given a shadow of their magnificent reality.

Complexity theory teaches that real systems have other descriptions that are equally valid along with the classical, reductionist analysis. Network thermodynamics allows us to see some concrete example of what this means. By unifying the non-mechanistic thinking characteristic of thermodynamic reasoning with specific mechanistic realizations of very complicated systems it allows us to make the necessary transition from formalisms centered on mechanism to the complimentary formalisms which forsake that mechanistic detail for a way of understanding self-referential context dependent relationships which are at the heart of the real system. Network thermodynamics along with all the other mechanistic tools used in complexity research can never go beyond the scope of the largest model of reductionism, the dynamic system.

This introduction to the formal aspects of Network Thermodynamics has given a few examples of ways in which understanding of systems can be enhanced using Network Thermodynamics as an example of combining some non-mechanistic ideas with specific mechanistic models. The use of Network Thermodynamics was deliberate, but other ways of combining these two distinct ways of looking at systems are now well known in complexity research. What is almost universally ignored is that all these mechanistic approaches do not really deal with system complexity, rather they are a way of handling complication.

5.4. Conclusion: there is no conclusion

Network Thermodynamics has been demonstrated to furnish a list of parts that can be combined using specific topologies to model the mechanistic aspects of the organism. There is no limit to the size or level of complication of the system that this formalism can handle at the level of this mechanistic modeling. This is not enough.

Parts plus topology are not an organism

Have the necessary parts and topological schemes been discovered to construct or fabricate a mechanistic model of the organism? The Network Thermodynamic modeling formalism provides one way of seeing the role of organization in determining specific functional mechanisms. The metabolism

of the organism in the relational systems theory formalism contains the following processes:

- Diffusion
- Electrical events
- Bulk flow
- Chemical reactions

Included in chemical reactions are the code carrying reaction processes involving DNA and RNA leading to protein synthesis and, in particular, the synthesis of the very protein enzymes that allow the DNA and RNA to be of any use at all. This should be all that is necessary. But each network type listed here has its own character. The network formalism provides a common mathematical formalism to solve the problem of turning the network into a dynamic system of the traditional kind. The analogy between this formalism and the particle dynamics example given earlier is striking.

The use of n-ports allows larger networks combining all four of these physical processes into a larger, more complete mechanistic model. Different topological arrangements of a myriad of parts assembled into n-ports yields the many functional mechanisms in the organism.

The topology of bulk flow networks corresponds to the network of duct, blood vessels, cell membranes, etc that act as pathways for this flow. Electrical events are associated with neural networks, cell membranes, etc. Diffusion is organized into intricate networks by the sophisticated compartmentalization due to cellular structure and the organelles in the cells.

The chemistry is a world of its own yet linked to these other networks via n-ports that reflect the need for reactants and products to diffuse or be carried by bulk flow as well as the electrochemistry associated with charged molecules. The chemical events themselves are organized into elaborate networks of reactions. It would seem that, in principle at least, there are enough tools and parts and organizational schemes in this mechanistic shadow of the real organism to put everything together and create a model of the organism.

This is *not* enough. The functional components are so abstract and also so basic to the need for the closure of efficient cause that the system does not lend itself to “reverse engineering”. In other words, understand physiology is not adequate if what is desired is an understanding of how the organism can be fabricated.

This has been known for a long time and has been simply stated. The Cell Theory has observed that cells always come from other cells. More recent acknowledgements of this idea coined the term “autopoiesis” (Maturana and Varela [31]).

These concepts are difficult for science to deal with because the quest for knowledge drives investigation with a belief in the possibility of finding answers to all well posed questions. The issue then may be in the posing of the question (Mikulecky [24]). A study such as this one that suggests that the synthesis of a cell is beyond the scope of the reductionist’s mechanistic formalism is certainly open to question at this point in history.

There is an interesting and completely linked parallel between this question and that of the computability of complex systems beyond their mechanistic shadows. In both cases last resort answers to the evidence that there are real limits revolve around a hope that someday a point will be reached where the apparent obstacles to understanding will be overcome by the advancement of our knowledge through investigation. One can speculate that the same attitude must have been present when perpetual motion machines were the objective of the scientific quest. Thermodynamics put that idea to rest once and for all. Relational systems theory has provided another, related non-mechanistic formalism to shed light on still one more aspect of complex reality. The context dependence and self-reference inherent in all real systems make the use of this formalism necessary if an understanding of why the mechanistic approach has failed to go beyond its self imposed limits is to be understood in context. Only by moving to a level of abstraction that divorces the parts and their topology from the observed action of the context dependent functional components as was done using the [M, R] systems and the relational formalism can the question of why organisms are different from machines and why they can not be constructed from mechanistic models can be answered.

This does not parallel the situation that occurred with perpetual motion machines in every way. In that case the synthesis of the desired mechanism was proven to be impossible. In the case of the organism, the system exists in the real world, but the ability to fabricate it is demonstrated to be outside the realm of mechanistic models. This does not mean that organisms cannot be constructed. It directs attention to the need for a new paradigm for fabrication that is not based on reverse engineering and mechanistic models. The relational model having given the insight that the desired system must be closed to efficient cause also falls short of the mark when we ask how this

can be synthesized. It was formulated to answer the question “why?” rather than “how?”. The use of entailment in answering the why question does show another insight. The question becomes one of transferring entailment from the creator of the system to the system itself. This concept seems entirely outside the present body of our knowledge. More is needed. In fact, the key to the fabrication problem is missing.

The self—referential circle is endless it seems. Hundreds of years of successful scientific inquiry has produced a technological world worth the awe it engenders. Even though it has not been mentioned specifically here, this, of course, includes quantum mechanics. Our quest for understanding of the universe in the large and the sub-atomic world in the small grows by opening new questions rather than by bringing and end.

Finally, the acknowledgement of the subjective aspects of our perception and formation of the models we use to view the real world makes it possible to reach a new level of understanding about the quest for knowledge. It enables us to step outside of the bounds traditional science has constructed in its attempts to eliminate this subjectivity and to view reality in a fresh way that encompasses the old and the new. The modeling relation applies to science as a world model as it does to all models our minds construct. In particular, the kind of abstraction used in constructing the [M, R] systems model can be applied to the modeling relation itself. The making of models of the modeling process is certainly context dependent as well as self-referential.

Where does this all lead? Once science is seen as a world model or belief structure using the inbuilt subjectivity of the modeling relation, the practice of imposing limits on models of reality and conforming to its largest model is open to scrutiny just like any other belief structure. The comparison of science with other belief structures becomes an obvious next step (Kercel and Mikulecky [144]).

Is this not what complex reality demands for its understanding? The definition of “complex” that lies behind all the concepts discussed here is that complex reality demands that there can be no largest model. Instead there must be a multitude of models corresponding to the multitude of distinct ways for interacting with the system. “Distinct” as used in this context means that the models resulting from different ways of interacting with the system cannot be derived from each other. The number of ways of interacting with a system as intricate as the human mind and its interacting with other minds as well as with the material aspects of complex reality suggests that the number of models needed is indeed infinite.

The circle is indeed endless and the conclusion is the beginning. Each new understanding changes the context of the system in a self-referential way so that what was known is now different and this necessitates a new model to incorporate this. At this point it seems best to allow the cycle of knowing to go on without trying to elaborate further. This is the best that can be done from this perspective at the moment.

References

1. K. Popper, *Objective Knowledge*. Oxford University Press, Revised Edition (1979). First published: Oxford University Press (1972).
2. W. Dress, Epistemology and Rosen's Modeling Relation, [http:// HyperNews. ngdc. noaa.gov/ISSS/get/WILL.html](http://HyperNews.ngdc.noaa.gov/ISSS/get/WILL.html) (1999).
3. A. Katchalsky and PF Curran, *Non-Equilibrium Thermodynamics in Biophysics*, Harvard Univ Press, Cambridge, MA (1965).
4. L. Peusner, *The Principles of Network Thermodynamics and Biophysical Applications*, Ph D Thesis, Harvard. Univ., Cambridge, MA. (1970), [Reprinted by Entropy Limited, South Great Road, Lincoln, MA 01773 (1987)].
5. L. Peusner, *J Chem Phys* **77**, 5500-5507 (1982).
6. L. Peusner, Electrical Network Representation of N-Dimensional Chemical Manifolds, in *Chemical Applications of Topology and Graph Theory*, R. B. King, (ED.), Elsevier, Amsterdam (1983).
7. L. Peusner, *J. Theor. Biol.* **102**, 7-39 (1983).
8. L. Peusner, *J. Theor. Biol.* **115**, 319-335 (1985).
9. L. Peusner, *J. Chem. Phys.* **83**, 1276-1291 (1985).
10. L. Peusner, *J. Chem. Soc. Faraday. Trans. 2*, **81**, 1151-1161 (1985).
11. L. Peusner, *J. Theor. Biol.* **122**, 125-155 (1986).
- 12.
13. L. Peusner, *Studies in Network Thermodynamics*, Elsevier, Amsterdam (1986).
14. G. F. Oster, A. Perelson, and A. Katchalsky, *Quart Rev Biophys* **6**, 1-134 (1973).
15. E. D. Schneider and J. J. Kay, *Mathematical and Computer Modelling* **19**, 25-48 (1994).
16. R. Rosen, *Anticipatory Systems: Philosophical, Mathematical & Methodological Foundations*, New York, Pergamon Press (1985).
17. R. Rosen, Some Relational Cell Models: The Metabolism-Repair System, in *Foundations of Mathematical Biology*, Vol. 2, Academic Press, N.Y. & London (1972) p. 217-253.
18. R. Rosen, *Fundamentals of Measurement and Representation of Natural Systems*. North-Holland, New York (1978).
19. R. Rosen, *Anticipatory Systems: Philosophical, Mathematical, & Methodological Foundations*. Pergamon Press, New York (1985).

20. R. Rosen, *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. Columbia University Press, New York (1991).
21. R. Rosen, On the Limits of Scientific Knowledge, in *Boundaries and Barriers: On the Limits to Scientific Knowledge*, JL Casti and A Karlqvist (eds.), Addison-Wesley, Reading (1996) pp. 199-214.
22. R. Rosen, *Essays on Life Itself*. Columbia University Press, New York (2000).
23. D. C. Mikulecky, *Comput Chem* **25**, 317-328 (2001).
24. D. C. Mikulecky, *Acta Biotheoretica* **44**, 179-208 (1996).
25. D. C. Mikulecky and Robert Rosen *Syst Res Behav Sci* **45**, 419-432 2000. D. C Mikulecky, Robert Rosen, in: *Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 9; *Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Data Mining, and Complex Systems*, C. H. Dagli, A. L. Buczak, J. Ghosh, M. J. Embrechts, and O. Ersoy, (eds.), ASME Press, NY (1999) pp. 193-198.
26. D. C. Mikulecky, *Comput Chem* **25**, 341-348 (2001).
27. J. Horgan, *The End of Science: Facing the Limits of Science in the Twilight of the Scientific Age*, Broadway Books, New York (1996).
28. D. D. Hoffman, *Visual Intelligence: How We Create What We See*, WW Norton & Co, New York (1998).
29. J. Horgan, *Sci. Amer.*, **June**, 104-109 (1995).
30. N. Rashevsky, *Bull. Math. Biophys.* **16**, 317-349 (1954).
31. H. R. Maturana and F. J. Varela. *Autopoieses and Cognition: The Realization of the Living*. D. Reidel, Dordrecht (1980).
32. L. O. Chua and R. N. Madan, *IEEE Circuits and Devices Magazine* **4**, 3-13 (1988).
33. J. Meixner, *J. Math. Phys.* **4**, 154-159 (1963).
34. J. Meixner, Network Theory in its Relation to Thermodynamics, in *Proceedings of the Symposium on Generalized Networks*, J. Fox, ed., Wiley Interscience, NY, p. 13-25, 1966.
35. G. F. Oster, A. Perelson, and A. Katchalsky, *Nature* **234**, 393-399 (1971). (See editorial: "Networks in Nature, pp. 380-381, same issue).
36. G. F. Oster and C. A. Desoer, *J. Theor. Biol.* **32**, 219-241 (1971).
37. G. F. Oster and A. S. Perelson, *Israel J. Chem.* **11**, 445-478 (1973).
38. G. F. Oster and D. M. Auslander, *J. Franklin. Inst.* **292**, 1-13 (1971).
39. G. F. Oster and D. M. Auslander, *J. Franklin. Inst.* **293**, 77-90 (1971).
40. G. F. Oster and A. S. Perelson, *Arch. Rational Mech. Anal.* **55**, 230-274 (1974).
41. P. Penfield, Jr, R. Spence, and S. Duinker, *Tellegen's Theorem and Electrical Networks*; Research Mon.# 58, M.I.T. Press, Cambridge, MA (1970).
42. S. Perelson, *Biophys. J.* **15**, 667-685 (1975).
43. S.. Perelson and G. F. Oster, *Arch. Rational Mech. Anal.* **57**, 31-98 (1974).
44. S. Perelson, Toward a Realistic Model of the Immune System, in *Theoretical Immunology*, AS Perelson (ed.), Addison-Wesley, Redwood City, CA (1988) pp. 377-401.

45. A. Desoer and G. F. Oster, *Int. J. Eng. Sci.* **11**, 141-155 (1973).
46. C. Mikulecky and FA Sauer, *J. Math. Chem.* **2**, 171-196 (1988).
47. C. Mikulecky, *Applications of Network Thermodynamics to Problems in Biomedical Engineering*. New York University Press: New York (1993).
48. W. A. Blackwell, *Mathematical Modeling of Physical Networks*, Macmillan, NY (1968).
49. P. C. Breedveldt, *Physical Systems Theory in Terms of Bond Graphs*, Ph.D. Thesis, Enschede, The Netherlands (1984).
50. V. D. Gebben, *J. Franklin Inst.* **308**, 361-369 (1979).
51. Karnopp and R. C. Rosenberg, *Analysis and Simulation of Multiport Systems: The Bond Graph Approach to Physical Systems*, M.I.T. Press, Cambridge, MA (1968).
52. D. Karnopp and R. C. Rosenberg, *System Dynamics: A Unified Approach*, Wiley, NY (1975).
53. H. E. Koenig, Y. Tokad, and H. K. Kesevan, *Analysis of Discrete Physical Systems*, McGraw-Hill, NY (1967).
54. G. J. MacFarlane, *Dynamic System Models*, Harrap, London (1970).
55. V. C. Rideout, *Mathematical and Computer Modeling of Physiological Systems*, Prentice Hall, Englewood Cliffs, NJ (1991).
56. P. H. Roe, *Networks and Systems*, Addison-Wesley, Reading, MA (1966).
57. J. U. Thoma, *Introduction to Bond Graphs and Their Applications*, Pergamon, NY (1975).
58. J. L. Wyatt, *Comp. Prog. Biomed.* **8**, 180-195 (1978).
59. J. L. Wyatt, D. C. Mikulecky, and J. A. De Simone, *Chem. Eng. Sci.* **35**, 2115-2128 (1980).
60. D. C. Mikulecky, The Use of a Circuit Simulation Program (SPICE2) to Model the Microcirculation, in *Biofluid Mechanics* Vol. 2, D. J. Schneck (ed.) (1980) p. 327-345.
61. D. C. Mikulecky, A Network Thermodynamic Approach to the Hill-King & Altman Approach to Kinetics: Computer Simulation, in *Membrane Biophysics II: Physical Methods in the Study of Epithelia*, M. Dinno, A. B. Calahan, and T. C. Rozzell (eds.), AR Liss, NY (1983) pp. 257-282.
62. D. C. Mikulecky, *Math. Biosci.* **72**, 157-179 (1984).
63. D. C. Mikulecky, Network Thermodynamics in Biology and Ecology: An Introduction, in: *Ecosystem Theory for Biological Oceanography*, R. E. Ulanowicz and T. Platt (eds.), *Canadian Bull. Fisheries and Aq. Sci.* **231**, DC 163-175 1985.
64. D. C. Mikulecky, Topological Contributions to the Chemistry of Living Systems. in *Graph Theory and Topology in Chemistry*, RB King and DH Rouvray (eds.), Elsevier, NY (1987) p. 115-123.
65. D. C. Mikulecky, Network Thermodynamics: A Unifying Approach to Dynamic Non-linear Living Systems, in *Theoretical Ecosystems Ecology: The Network Perspective*, TP Burns and M Higashi (eds.), Cambridge Univ. Press (1991) pp. 71-100.

66. D. C. Mikulecky, E. G. Huf, and S. R. Thomas, *Biophys. J.* **25**, 87-105 (1979).
67. Mintz, S. R. Thomas, and D. C. Mikulecky, *J. Theor. Biol.* **123**, 1-19 (1986).
68. E. Mintz, S. R. Thomas, and D. C. Mikulecky, *J. Theor. Biol.* **123**, 21-34 (1986).
69. L. Peusner, D. C. Mikulecky, S. R. Caplan, and B. Bunow, *J. Chem. Phys.* **83**, 5559-5566 (1985).
70. D. E. Oken, S. R. Thomas, and D. C. Mikulecky, *Kidney Int.* **19**, 359-373 (1981).
71. R. L. Seither, D. F. Trent, D. C. Mikulecky, T. J. Rape, and I. D. Goldman, *J. Biol. Chem.* **264**, 17016-17023 (1989).
72. R. L. Seither, D. F. Trent, D. C. Mikulecky, T. J. Rape and I. D. Goldman, *J. Biol. Chem.* **266**, 4112-4118 (1991).
73. R. L. Seither, D. Hearne, D. Trent, D. C. Mikulecky, and I. D. Goldman, *Computers Math. Appl.* **20**, 87-101 (1990).
74. D. B. Talley, J. P. Ornato, and A. M. Clarke, *Biomed. Inst. Tech.* **July/August**, 283-288 (1990).
75. K. M. Thakker, J. H. Wood, and D. C. Mikulecky, *Comp. Prog. Biomed.* **15**, 61-72 (1982).
76. K. M. Thakker and D. C. Mikulecky, *Math. Modeling.* **7**, 1181-1186 (1985).
77. D. Walz, *Biochim. Biophys. Acta* **1019**, 171-224 (1990).
78. D. Walz, S. R. Caplan, D. R. L. Scriven, and D. C. Mikulecky, *Methods of Bioelectrochemical Modeling*, in *Treatise on Bioelectrochemistry*, G. Millazzo (ed.), Birkhauser (1992).
79. J. C. White, *J. Biol. Chem.* **254**, 10889-10895 (1979).
80. J. C. White, *Bull. Math. Biol.* **48**, 353-380 (1986).
81. J. C. White and D. C. Mikulecky, *Pharmacol. Ther.* **15**, 251-291 (1981).
82. M. B. Cable, J. J. Feher, and F. N. Briggs, *Biochem.* **24**, 5612-5619 (1985).
83. J. J. Feher, *J. Biol. Chem.* **257**, 10191-10199 (1982).
84. J. J. Feher, C. S. Fullmer, and R. H. Wasserman, *Amer. J. Physiol.* **262**, C517-C526 (1992).
85. M. L. Fidelman and S. Mierson, *Am. J. Physiol.* **257**, G475-G487 (1989).
86. S. Mierson and M. L. Fidelman, *The Role of Epithelial Ion Transport in Taste Transduction: A Network Thermodynamic Model*, in *Advances in Mathematics and Computers in Medicine*, Vol. 6, D. C. Mikulecky and M. Witten (eds.), Pergamon Press (1992) p. 119-134.
87. M. Fidelman and D. C. Mikulecky, *Am. J. Physiol.* **250**, C978-C991 (1986).
88. M. L. Fidelman and D. C. Mikulecky, *J. Theor. Biol.* **130**, 73-93 (1988).
89. P. Cruziat and R. Thomas, *Agronomie* **8**, 613-623., 1988.
90. L. J. Goldstein and E. B. Rypins, *Comp. Methods. Prog. Biomed.* **29**, 161-172 (1989).
91. J. C. Horno, F. Gonzalez-Fernandez, A. Hayas, and F. Gonzalez-Caballero, *Biophys. J.* **55**, 527-535 (1989).

92. J. C. Horno, F. Gonzalez-Fernandez, A. Hayas, and F. Gonzalez-Caballero, *J. Memb. Sci.* **42**, 1-12 (1989).
93. E. G. Huf and J. R. Howell, *J. Memb. Biol.* **15**, 47-66 (1974).
94. E. G. Huf and D. C. Mikulecky, *J. Theor. Biol.* **112**, 193-220 (1985).
95. E. G. Huf and D. C. Mikulecky, *Am. J. Physiol.* **250**, F1107-F1118 (1986).
96. J. M. May and D. C. Mikulecky, *J. Biol. Chem.* **258**, 4771-4777 (1983).
97. J. M. May and D. C. Mikulecky, *J. Biol. Chem.* **257**, 11601-11608 (1982).
98. D. C. Mikulecky and M. Thellier, *C. R. Acad. Sci. III* **316**, 1399-1403 (1993).
99. J. Prideaux, *Acta Biotheoretica* **44**, 219-233 (1996).
100. R. Abraham and C. D. Shaw, *Dynamics: The Geometry of Behavior; Part 1, Periodic Behavior*, Aerial Press, Santa Cruz, CA (1982).
101. R. Abraham and C. D. Shaw, *Dynamics: The Geometry of Behavior; Part 2, Chaotic Behavior*, Aerial Press, Santa Cruz, CA (1983).
102. R. Abraham and C. D. Shaw, *Dynamics: The Geometry of Behavior; Part 3, Global Behavior*, Aerial Press, Santa Cruz, CA (1984).
103. R. Abraham and C. D. Shaw, *Dynamics: The Geometry of Behavior; Part 4, Bifurcation Behavior*, Aerial Press, Santa Cruz, CA (1988).
104. R. Abraham, R. and C. D. Shaw, *Dynamics: A Visual Introduction*, in *Self-Organizing Systems: The Emergence of Order*, FE Yates (ed.), Plenum Press, NY (1987).
105. M. Arbib and E. G. Manes, *Arrows, Structures, and Functors: The Categorical Imperative*, Academic Press, New York (1975) pp. 93-106.
106. R. Abraham and J. E. Marsden, *Foundations of Mechanics*, Benjamin/Cummings, Reading, MA (1978).
107. H. Branin Jr, The Algebraic-Topological Basis for Network Analogies and the Vector Calculus, in: *Proceedings of the Symposium on Generalized Networks*, J Fox, Ed, Polytechnic Press, Brooklyn, NY (1966) pp. 453-491.
108. D. C. Mikulecky, *Comput. Chem.* **25**, 369-392 (2001).
109. D. H. Tellegen, *Phillips Res. Rep.* **7**, 259-269 (1952).
110. B. Callen, *Thermodynamics*, Wiley, NY (1960).
111. G. N. Hatsapoulos and J. H. Keenan, *Principles of General Thermodynamics*, Wiley, NY (1965).
112. Prigogine and R. Defay, *Chemical Thermodynamics*, Longmans Green, London (1965).
113. L. Tisza, *Generalized Thermodynamics*, MIT Press, Cambridge, MA (1977).
114. Truesdell, *Rational Thermodynamics*, McGraw-Hill, NY (1969).
115. S. R. deGroot and P. Mazur, *Non-Equilibrium Thermodynamics*, North-Holland, Amsterdam (1962).
116. Fitts, *Non-Equilibrium Thermodynamics*, McGraw-Hill, NY (1962).
117. I. Prigogine, *Thermodynamics of Irreversible Processes*, Wiley, NY (1961).
118. S. R. Caplan and A. Essig, *Bioenergetics and Linear Non-Equilibrium Thermodynamics: The Steady State*, Harvard, Cambridge, MA (1983).

119. C. Mikulecky, WA Wiegand, and JS Shiner, *J. Theoret. Biol.* **69**, 471-510 (1977).
120. L. O. Chua, *IEEE Trans. Cir. Theory* **CT-18**, 507-519 (1961).
121. R. Kirchhoff, On the Solution of the Equations Obtained from the Investigation of the Linear Distribution of Galvanic Currents, 1847, English translation in *Graph. Theory 1736-1936*, N. L. Biggs, E. K. Lloyd, and R. J. Wilson (eds.), Oxford (1976).
122. D. C. Mikulecky and S. R. Thomas, *J. Franklin Inst.* **308**, 309-325 (1979).
123. P. W. Tuinenga, *SPICE: A Guide to Circuit Simulation and Analysis Using PSPICE*. Prentice Hall, NJ (1988).
124. P. M. DeRusso, R. J. Roy, and C. M. Close, *State Variables for Engineers*, Wiley, NY (1965).
125. L. O. Chua, *Introduction to Nonlinear Network Theory*, McGraw-Hill, NY (1969).
126. L. O. Chua and P. Lin, *Computer -Aided Analysis of Electronic Circuits: Algorithms and Computational Techniques*, Prentice-Hall, Englewood Cliffs, NJ (1975).
127. L. O. Chua and Y. Lam, *IEEE Trans. Cir. Theory* **CT-20**, 370-381 (1973).
128. D. C. Mikulecky, *Math. Comp. Modeling* **19**, 99-118 (1994).
129. S. R. Thomas and D. C. Mikulecky, *Am. J. Physiol.* **235**, F638-F648 (1978).
130. L. Onsager, *Phys. Rev.* **37**, 405-426 (1931).
131. Onsager, *Phys. Rev.* **38**, 2265-2279 (1931).
132. S. J. Mason and H. J. Zimmermann, *Electronic Circuits, Signals, and Systems*, Wiley, NY (1960).
133. Kedem and A. Katchalsky, *Trans. Faraday Soc.* **59**, 1918-1930 (1963).
134. O. Kedem and A. Katchalsky, *Trans. Faraday Soc.* **59**, 1931-1940 (1963).
135. O. Kedem and A. Katchalsky, *Trans. Faraday Soc.* **59**, 1941-1953 (1963).
136. B. Callen, Asymmetry Interpretation of Thermodynamics, in *Foundations of Continuum Thermodynamics*, J. J. D. Domingos, M. N. R. Nina, and J. H. Whitelaw (eds.), Wiley, NY (1973) pp. 61-79. (See comment by L. Tisza, p 79).
137. O. Kedem and S. R. Caplan, *Trans. Faraday Soc.* **61**, 1897-1911 (1965).
138. D. G. Miller, *Chem. Revs.* **60**, 15 (1960).
139. D. G. Miller, The Experimental Verification of the Onsager Reciprocal Relations, in *Transport Phenomena in Fluids*, H. J. M. Hanley (ed.), Dekker, NY (1969) pp. 377-432.
140. O. Kedem and A. Katchalsky, *Bioch. Biophys. Acta* **27**, 229-246 (1958).
141. O. Kedem and A. Katchalsky, *J. Gen. Physio.* **45**, 143-179 (1961).
142. A. DeSimone and S. R. Caplan, *Biochem* **12**, 3032-3039 (1973).
143. A. DeSimone and S. R. Caplan, *J. Theo. Biol.* **39**, 523-544 (1973).
144. S. W. Kercel and D. C. Mikulecky, Why do people behave religiously? *Evolution and Cognition*, in press.
145. J. Barwise and L. Moss, *Vicious Circles*, CSLI Publications, Stanford, CA (1996).

Chapter 4

GRAPHS AS MODELS OF LARGE-SCALE BIOCHEMICAL ORGANIZATION

Pau Fernández

ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80,
08003 Barcelona, Spain

Ricard V. Solé

ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80,
08003 Barcelona, Spain

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA

1. Introduction

Cells have many different types of molecules interacting in space and time to produce the coordinated behavior we observe. They are a beautiful example of complex systems, which are difficult to study in their entirety due to the overwhelming number of components they have. However, allowing for a certain degree of simplification, one can always start looking at these systems by simply measuring what pairs of units in these multicomponent systems engage in some form of interaction. This crude, discrete information can be a very informative first step towards an understanding of cells as an integrated whole, and provides the most simple global picture one can obtain of these systems: their underlying network.

Networks pervade biology, and they are present at many different spatial and temporal scales, from molecular biology to large-scale evolution. By understanding the possible scenarios responsible for the emergence of their architecture, great insight can be gained on their evolutionary origins [1]. As a first example, we can consider the case of protein molecules. On the one hand, they can be viewed simply as the units on top of which a complex system like the cell is built. But on the other hand, they count as complex systems in their own right, given the complexity of their structure and folding process. A three-dimensional structure of a linear biopolymer like a protein can, in fact, be roughly described by its contact structure, that is, by the list of all pairs of monomers that are spatial neighbors, given some

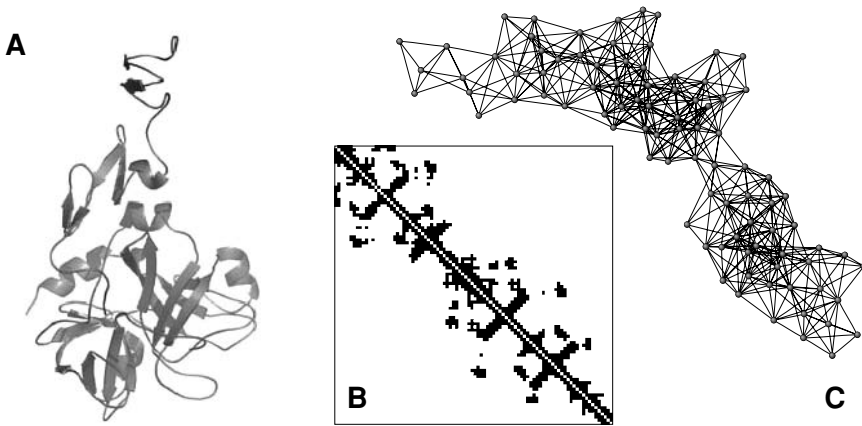


Figure 4.1. (A) A fragment of the structure for the human blood coagulation factor Xa (pdb 1xka); (B) The associated contact matrix. Black dots represent above threshold proximity between alpha carbons in the sequence, which are arranged as rows and columns in the order of the backbone; (C) The associated graph in which nodes represent amino acids, and links represent contact between them.

cut-off distance. Figure 4.1A shows the structure of a portion of a protein, and Figure 4.1B shows its contact matrix, the simplest way to quickly view the contact pairs. In this matrix, rows and columns represent each amino acid in the sequence, and a square is drawn black when the corresponding amino acids are in close proximity in the three dimensional structure.

Using this matrix, one can actually construct a network, shown in Figure 4.1C, in which the vertices (or nodes) represent amino acids, and an edge (or link) is present whenever these two amino acids are in contact. A network such as this, as we will see throughout the chapter, can give us many hints about the system it is representing. To start with, the network in Figure 4.1C tells us that this protein is modular. Although we already see that this protein is modular from the examination of its structure, the important point to emphasize is that the network alone mirrors this modularity, and this is especially relevant whenever we don't have a clear picture of the system for which the network was measured.

Many other properties are in fact measurable just from the networks, which describe other equally interesting features. Applying various concepts which can be defined on graphs, we can examine the networks of

different systems trying to understand their overall features, and we can try to make sense of their connection patterns. In this chapter we will see how some of the properties of complex systems can be approached through the study of the properties of their corresponding networks, the so-called complex networks.

2. Basic Properties of Random Graphs

A graph G is defined by a set of N vertices (or nodes) $V = \{v_1, v_2, \dots, v_N\}$ and a set of L edges (or links) $E = \{e_1, e_2, \dots, e_L\}$, which connect pairs of vertices. Depending on the existence of directionality in the edges, graphs can be either directed or undirected. Figure 4.2 depicts two graphs, one directed and one undirected, using arrows as a representation of directed edges. A certain graph G can, by this simple definition, represent the structure of a given system, in which, nodes correspond to units and edges to interactions. In the following, we will be concerned with graphs in which edges have no weights attached, although many interesting results exist for weighted graphs.

The number of edges k that arrive to vertex v is called degree, and it is divided, in the case of directed graphs into k_i , the in-degree, and k_o ,

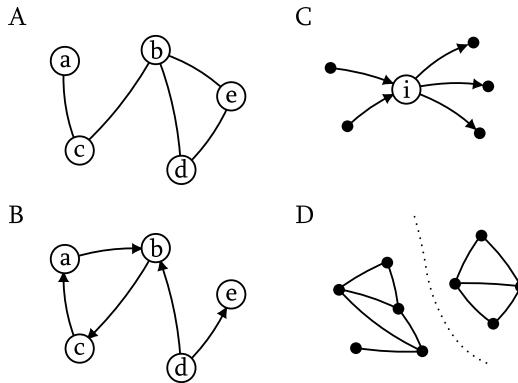


Figure 4.2. (A) An example of undirected graph G with $V = \{a,b,c,d,e\}$ and $E = \{(a,c), (b,c), (b,d), (b,e), (d,e)\}$; (B) Example of a directed network with the same V and $E = \{(a,b), (b,c), (c,a), (d,b), (d,e)\}$; (C) A node with degree $k = 5$ and indegree $k_i = 2$ and outdegree $k_o = 3$; (D) An undirected graph with two components.

the out-degree, with $k = k_i + k_o$, as shown in Figure 4.2C. The average degree is, then, $z = \langle k \rangle = 2L/N$, since every edge is attached to two different vertices. The distance between two vertices in the graph $d(v_i, v_j)$ is defined as the shortest number of links that have to be crossed to reach v_j from v_i , and since we are not dealing with weighted edges, it is an integer number. As an example, in the graph of Figure 4.2A, $d(a, e) = 3$ and $d(c, d) = 2$.

A random graph is a graph in which the pairs of connected vertices have been selected uniformly at random between all the possible pairs of vertices. Basically, this is equivalent to consider that every edge in the graph is present with an independent probability p , and absent with probability $1 - p$. In fact, there is a whole ensemble of graphs for a single value of p , and random graph theory deals with the average properties of these ensembles. The average number of edges in these graphs is thus $\langle L \rangle = \frac{N(N-1)}{2}p$ and accordingly, the average degree is, then $z = \frac{2\langle L \rangle}{N} \approx Np$, the last equality being valid only for large N . This kind of random graph were first studied by Erdős and Rényi, and many of their average properties are have been solved analytically [2] in the limit for large N . Random graphs are important because they represent the null hypothesis about systems that can be modeled as graphs. The study of their properties is thus crucial to understand real graphs, since the comparison between random graphs and real ones tells us what is relevant about the latter.

2.1. Degree distribution

One of the first interesting aggregate measures to consider is the degree distribution, $P(k)$ or p_k . This function describes the graph by giving the fraction of nodes that have a certain degree k . As an example, we can calculate p_k for a random graph, given the probability of linking two of its vertices, p , which is

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{z^k e^{-z}}{k!}. \quad (2.1)$$

The second equality becomes exact in the limit of large N , when a finite average degree $z = p/N$ makes p necessarily very small. This is, in fact, the Poisson distribution, an important fact to remember, as we will see.

2.2. Components

Depending on the exact organization of the L edges, a certain graph can have different components. Components are defined as the subsets of V in which all vertices of the subset can be reached from other members of the subset, but not from any other vertices. Figure 4.2D depicts a graph with two components. In random graphs, components have an important role. Basically, as z grows, a random graph shows a phase transition at which the so-called giant component forms.

For small values of z , the few edges present in the graph are unable to connect the vertices, and the result is a graph with many small components, having an average size that remains constant with larger N . However, for a critical value of z , a given fraction α of vertices form a single component, therefore having a size αN which scales linearly as N grows. This is the giant component, and it goes with all the rest of smaller components, which still remain constant in size as the graph grows. The first appearance of the giant component occurs, in fact, at $z = 1$.

2.3. Average path length

Another important average quantity in graphs is the **average path length**,

$$\ell = \frac{1}{N(N-1)} \sum_{\forall i,j} d(v_i, v_j) \quad (2.2)$$

which actually represents the average distance between any pair of vertices. For graphs with given degree distribution, we can calculate ℓ in the following way [3]. Let's consider a certain vertex v , such as the one shown in Figure 4.3. First, we can calculate the average number of first neighbors, z_1 . This is, in fact, the same as z , since the degree measures the number of neighbors of a vertex. Now we can calculate the number of second neighbors, z_2 . To do this, we first need the distribution p'_k of number of remaining edges of a certain vertex, given that we arrive from one of them, chosen at random.

First, we must take into account the fact that a vertex with more degree has a greater probability of being accessed, i.e., proportional to its degree. Second, we also have to take into account the edge we arrived along, and

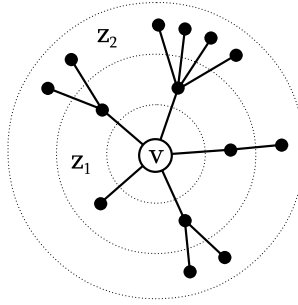


Figure 4.3. A vertex v surrounded by its first neighbors, on average z_1 , and its second neighbors, on average z_2 .

remove it from the counting. The result is, then,

$$p'_{k-1} = \frac{kp_k}{\sum_{i=0}^N kp_i}, \quad (2.3)$$

including the proper normalization. On average, then, we will find that following a random edge, the number of remaining edges which we can follow is

$$\sum_{k=1}^{\infty} (k-1)p'_{k-1} = \frac{\sum_{k=0}^{\infty} (k-1)kp_k}{\sum_{i=0}^N kp_i} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \quad (2.4)$$

That tells us the average of second neighbors found by following one of our first neighbors, but if we follow an average of $k_1 = k$ first neighbors, we have the total number of second neighbors from vertex v , that is,

$$z_2 = \langle k^2 \rangle - \langle k \rangle. \quad (2.5)$$

For the case of a random graph, with a Poisson distributed p_k , we have $z_2 = \langle k^2 \rangle$, so the mean number of second neighbors in the special case of a random graph is just the mean number of first neighbors squared, which doesn't hold in general.

We can now extend this calculation to further neighbors. The distribution p'_k also tells us how many neighbors a second neighbor has, and in fact it is true for neighbors at distance m in general. In brief, the average number

of neighbors at distance m of v is

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} = \frac{z_2}{z_1} z_{m-1} = \left[\frac{z_2}{z_1} \right]^{m-1} z_1. \quad (2.6)$$

Depending of the factor z_2/z_1 , the number of accessible neighbors can be infinite, something which reminds us about the phase transition mentioned above. The critical point for this transition is then, $z_2 = z_1$, which in the case of random graphs translates to $z = 1$, as already mentioned. In general, though, we can derive an expression for the equation $z_2 - z_1 = 0$ making use of Eq. (2.5), which gives us

$$\langle k^2 \rangle - 2 \langle k \rangle = \sum_{k=0}^{\infty} k(k-2)p_k = 0. \quad (2.7)$$

To return to our previous discussion, we can now calculate the average path length ℓ making use of Eq. (2.6). Basically, at some m in Eq. (2.6) the number of accessible neighbors will be N , and that is precisely ℓ , since it is the average number of steps at which an average node reaches the whole graph. Taking logarithms for equation $N = z_\ell$ and reorganizing, we arrive at

$$\ell = \frac{\log(N/z_1)}{\log(z_2/z_1)} + 1. \quad (2.8)$$

For the special case of Erőds-Rényi random graph, in which $z_2 = z^2$, the expression reduces to $\ell = \log N / \log z$. This is an important result, in fact. It says that, in random graphs, ℓ grows rather slowly with the size of the graph, and that a big graph with $N = 1000$ and $z = 4$ will have $\ell \approx 5$.

2.4. Clustering

Graphs can also display structure at the local scale, and one of the easiest ways to describe it is by measuring of the average clustering coefficient, or C . This coefficient is a measure applied to a given vertex v_i , and it measures how far a vertex is of being part of a clique. A clique is basically a small graph with full connectivity; every vertex is connected to every other. It is clear that if a certain vertex v_i is part of a clique, the number of edges between its k_i neighbors must be $k_i(k_i - 1)/2$. If instead of that we find E_i edges, the corresponding fraction is precisely C_i , the clustering coefficient,

that is,

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2.9)$$

When C_i is average for the whole network, we have

$$C = \frac{1}{N} \sum_{i=1}^N C_i = \frac{1}{N} \sum_{i=1}^N \frac{2E_i}{k_i(k_i - 1)} \quad (2.10)$$

The interest in clustering came from social networks, in which networks of acquaintances usually display a high degree of clustering: your friends tend to be friends of each other. This is in contrast to random graphs in which clustering is very small. Actually in random graphs, since each vertex is active with probability p , the clustering is always the same, that is $C = p = \langle k \rangle / N$. So clustering is very, very small for large graphs.

2.5. Small-worlds

Now that we have seen both the average path length and the property of clustering, we can use them to assess if a certain graph is “small world”. The small-world effect was demonstrated in the 1960s by Stanley Milgram in a famous experiment involving letter passing. He gave some letters to some friends of his, asking them to pass the letters to a friend of theirs that they thought was closest to the recipient in the letter. Contrary to logic, the number of steps, or friendships, required to reach the recipient was rather low, in the order of 6 or 7. But how could this happen when social networks have a very high degree of clustering? Although the average path length for a random graph goes as $\log N / \log z$, in a very clustered network the average path length must be larger perforce.

For instance, let's consider the network in Figure 4.4A. It's a quite clustered network. In particular, the clustering coefficient C of this network is precisely $1/2$ since all nodes are equivalent and the 4 neighbors of each node have 3 of the 6 possible connections between them. Having just 40 nodes, the clustering of an equivalent random network would be $4/40 = 0.1$ so this network has 5 times more clustering than its random counterpart. But in an equivalent one-dimensional network with $N = 1000$ nodes, clustering would still be $1/2$ (since it is independent of size), although the corresponding clustering of a random network would drop to $4/1000 = 0.004$,

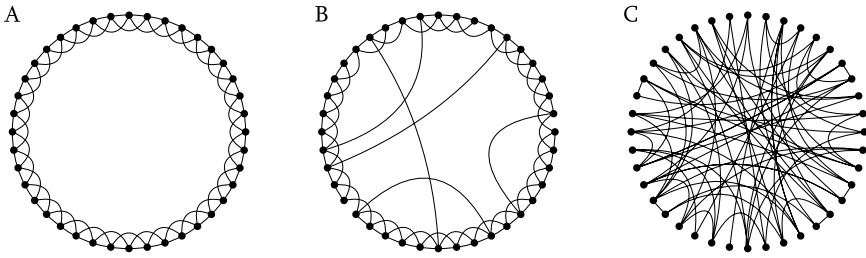


Figure 4.4. (A) An example of a strongly clustered network with the topology of a one dimensional space. To reach any vertex from any other, many links have to be crossed since all connections are local; (B) The same network as in (A) with some edges rewired at random; (C) A completely random network.

which makes a difference of three orders of magnitude. In general, we could think of this kind of clustered networks as a crude model of social networks which, when very large, have the relevant property of being highly clustered, in particular with respect to random networks.

But an important ingredient is missing. Although in a social network many of one's friends are friends themselves, there is a very small fraction of "long-distance" friends, removed from the local community of acquaintances that represent something similar to the "random" links of an Erősds-Rényi graph. As we will see, this small number of edges alone provides enough long-range jumps to make the average path length fall rapidly, and at the same time they don't affect the clustering very much, the so-called small-world effect.

The small-world effect was studied [4] using a model which considers precisely the network of Figure 4.4. Starting with it, a rewiring process is defined which takes each edge with probability p and rewires it to a random vertex. Figure 4.4B shows the initial graph with some random links rewired. When p approaches 1, the randomness of the graph is total, yielding a graph which resembles the one in Figure 4.4C. In this way, we have a graph which depends on p . The small world effect is clarified if we plot, as a function of p , the normalized clustering $C(p)/C(0)$ and the normalized average path length $L(p)/L(0)$, at the same time. The result is what is shown in Figure 4.5. As it is apparent, there is a broad region of p values (very small values, noting that the y axis is logarithmic) in which clustering is still high whereas the path length is already low. The

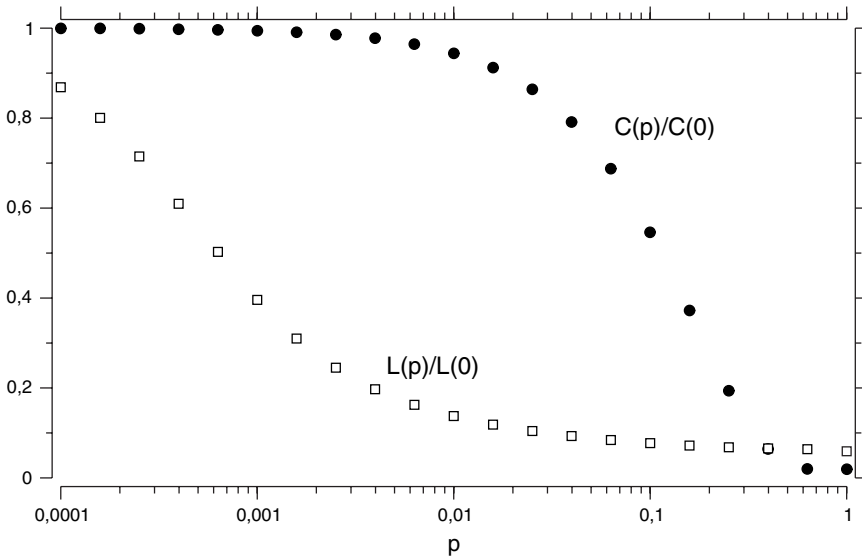


Figure 4.5. The small-world effect quantitatively. The graph shows the normalized clustering coefficient and the normalized average path length as a function of the probability of rewiring p . Note that the p axis has a logarithmic scale. For a very small value of p , around 0.01, the average path length has dropped substantially whereas the clustering coefficient is still almost untouched. The joint appearance of these two properties describe the small-world effect. The graphs used had $N = 1000$ and $\langle k \rangle = 10$, reproducing the results in [4].

explanation for this is found in the very small fraction of connections that connect remote parts of the system reducing the average path length to be comparable to that of a random graph.

3. Protein Structure and Contact Graphs

At the smaller scale, we can consider proteins as a key example of a complex system composed by many parts in interaction (the protein residues). As is well known from biochemistry and molecular biology, proteins play a key role in cell function and they are actually responsible for many different features of cell behavior, from shape to communication [5]. Here structure and function are intimately linked. From the linear chain of residues coded at the genome sequence level, the chain acquires a

functional shape by folding in three dimensions towards the so called native state. The final shape defines a higher-order structure in relation with the linear sequence and thus involves further, long-range interactions among residues.

Since the contact graph of amino acid interactions defines a complex network, we might first ask what type of overall pattern is found here. Protein architecture seems to display a number of features resulting from a selection process [6], and it is also clear now that folding is highly optimized in relation with what would be expected from a polymer exhibiting random interactions among residues. Using the previous tools of graph analysis, we can first explore the question of the presence of small world behavior in the protein contact graph.

3.1. Proteins are small worlds

As shown in Figure 4.6, the statistical pattern of organization of protein contact graphs reveals a well-defined small world topology. Clustering is typically much higher than expected from random wiring (for which we should observe $C \sim N^{-1}$) and the average path length scales with the logarithm of system's size N . It is interesting to see that we are actually dealing

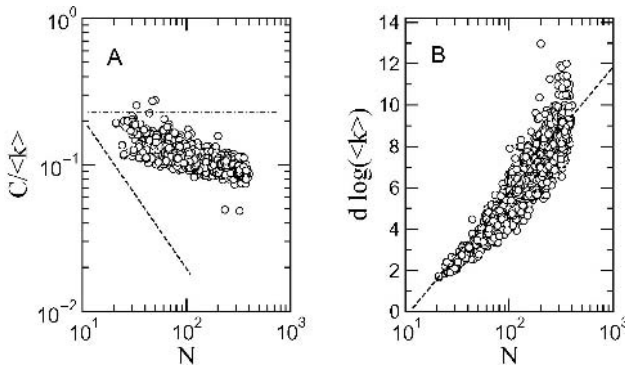


Figure 4.6. Small world structure of protein folding maps. Here the two key global measures are shown for a large number of proteins from the ArchDB database [7]. In (A) the clustering coefficient is shown (properly normalized with the average degree). The predicted scaling relation from a random graph (i. e. $C \sim N^{-1}$) is indicated as a dashed line. In (B) the average path length is plotted against size. The prediction from a small world architecture indicate that $d \sim \log N / \log z$. This is indicated as a dashed line, which is closely followed by the observed data set.

with a scenario of graph rewiring not so far from the Watts-Strogatz model. Starting with a linear chain of residues connected to only two nearest neighbors, protein folding involves the creation of a number of shortcuts which we can easily identify to those key residues responsible for the folding.

Beyond these global features, which indicate that some quantitative traits can be properly identified, other relevant properties can be measured, perhaps much closer to the functional characterization of protein graphs and other complex networks. Modular patterns can be observed by looking at how subsets of a given graph are connected among them. In its simplest terms, modules can be seen as groups of units which are more connected between them than with other parts of the graph. Modules have been found in biological systems at multiple levels, from RNA structures [8] to the cerebral cortex (see ref. 1). The widespread character of modular organization has been always associated to functionality, compartmentalization and evolution. The evolutionary conservation of modules is well known from different examples, particularly in early development [9-11]. The argument is that the special features of some of these modules are tightly linked to their robustness under different sources of noise. As such, they would be the units of selection.

The modular character of biological networks is assumed to be a consequence of both their robustness and evolvability [12,13]. In a different context, it has been suggested that modularity might arise from the intrinsic structure of the non-metric mapping between genotype and phenotype, at least in molecular networks [14]. Although functionality must influence the selection of some modular structures, we will see in Section 4.4 that proto-modules might actually emerge as inevitable patterns without any predefined functional meaning.

3.2. Hierarchical clustering in contact maps

Although there is no agreed measure of overall modularity for graphs, the intuitive notion that a module is a group of nodes that have higher connectivity to the inside than to the outside seems sensible, and some simple heuristic algorithms can reveal this straightforwardly [15]. To do so, the topological overlap (TO) of two nodes has to be defined. The degree of TO of two vertices tries to give a value in the range (0, 1) for the fact that two vertices belong to the same “module”. If they do, they will probably have many neighbors in common, and that is precisely what the TO value

measures. Its definition is

$$O(v_i, v_j) = O(v_j, v_i) = \frac{J(v_i, v_j)}{\min(k_i, k_j)} \quad (3.1)$$

where $J(v_i, v_j)$ denotes the number of nodes to which both v_i and v_j are linked, plus one if there is a direct link between them. The OT value is commutative, and can be seen as a measure of the proximity in terms of modules of two vertices. Given the matrix of OT values, we can apply the general algorithm for hierarchical clustering, which tries to group together those vertices in the system that have a high proximity, that is, high TO. Basically, this algorithm starts with a matrix with the original values, and repeatedly finds the highest value, grouping together the corresponding row and column (the two vertices) into one new aggregated vertex, and computes the new values for the pair grouped vertices, so as to be able to apply the same procedure to the resulting matrix until it collapses into a single value.

The result of this process is a tree, and also an ordering of the vertices of the graph. The tree results from the recursive grouping of nodes and it is a binary tree (one in which a branch has always two subbranches), as implied by the grouping procedure. The ordering results from the representation of the tree in a plane that gives a particular linear position of the leaves that represent the nodes. These two results enable us to display the topological overlap in a suitable way, as Figure 4.7 shows. On the left there is a graph with

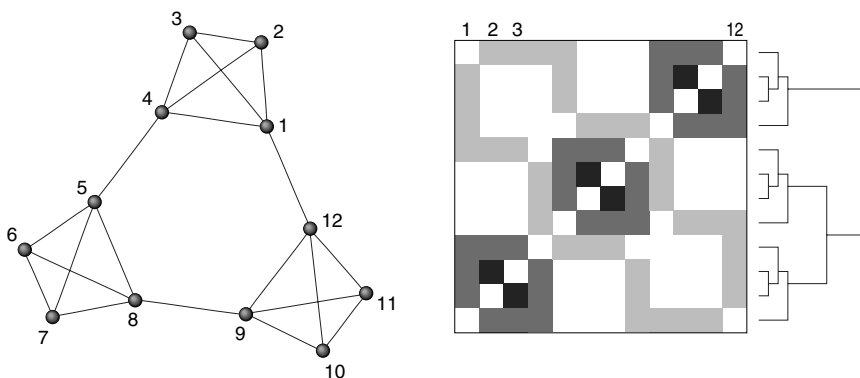


Figure 4.7. Left. A simple graph with 3 modules Right. The modular structure as viewed using hierarchical clustering on the topological overlap matrix.

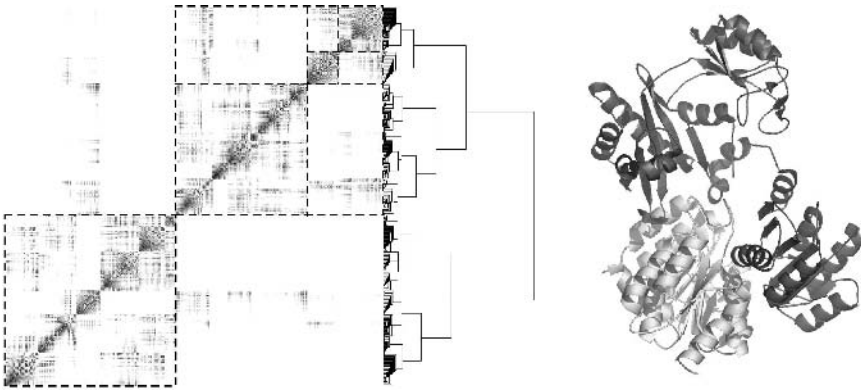


Figure 4.8. Left. Hierarchical organization of the Succinyl-CoA synthetase from the pig (*Sus scrofa*), as uncovered by the hierarchical clustering algorithm (see text). This method reveals two basic modules (corresponding to the two chains A and B), which are marked using a dashed box. These two modules also include a lot of internal substructure, shown by the further boxing of the upper right modules. Right. The corresponding protein structure (PDB code is 1euc).

evident modular structure, and on the right, the corresponding output of the hierarchical clustering algorithm. Values between 0 and 1 have been mapped to levels of gray. As shown, the indices in the graph correspond to the indices in the matrix, so for example, vertices 2 and 3 have a higher TO than others, and hence are in the central part of the modules, with a darker value. Modules in this diagram are organized along the darker regions of the diagonal, because mainly the algorithm has separated them in the linear ordering of the matrix. That is also apparent in the tree that appears on the right.

What is the pattern displayed by protein contact maps in terms of hierarchical clustering? An example is shown in Figure 4.8. We can clearly appreciate the presence of two well-defined modules, boxed using a dashed line. These modules can be identified using the tree, and also the fact that connections between them are much sparser than with the which are actually mapped into the two chains playing a functional role in this particular protein. An interesting feature that becomes obvious from the previous plot is that we actually have a rather complex, nested pattern of modularity: groups of proteins appear more connected at different scales and belong to

larger structures. In the figure, the box in the upper right part is also divided into two modules, of which the upper right one is further divided into two. Modularity is thus associated with hierarchical organization. This might actually be related with the fact that the process of folding is also hierarchical. The nested structure of the overlap map would be a fingerprint of the hierarchies involved in the folding process. Thus this method is able to identify the presence of well-defined domains in terms of topological arrangements, but can be used in the analysis of any other network structure. As we will see, modularity is actually a preeminent feature of the organization of complex networks.

4. Protein Interaction Networks

It is often said that the actions and properties of each cell are basically determined by the proteins it contains, which implement, like complex nanomachines, the tasks needed by the cell. We have already seen how these molecules are, in fact, usefully described in terms of their underlying graph. But in this constantly changing chemical world, proteins seldom work alone. By and large, almost all proteins are part of protein complexes, or at least engage in some form of interaction with other proteins. By means of physical contacts, proteins enable the cell to actively build structures, process signals from the environment, redirect chemicals to different metabolic pathways, or form the basis of gene regulation. A very useful picture of the organization of the cell can be obtained, therefore, from the information of which pairs of proteins interact with each other. By means of this information, a graph can be constructed, which describes the inner workings of the whole cell. However, the simple examination of the networks is not very useful, given their size. It is when we investigate this graph with the tools of network theory that some interesting properties become clearer.

In Figure 4.9A a part of the proteome network of *Homo sapiens* is shown. This network is one of the smallest in the DIP database. Although not especially useful as a detailed map, the network displays, nevertheless, many interesting properties. In comparison with Figure 4.9B, which is a random network with the same number of vertices and edges, the most important feature can readily be observed: the heterogeneity in the degree of vertices.

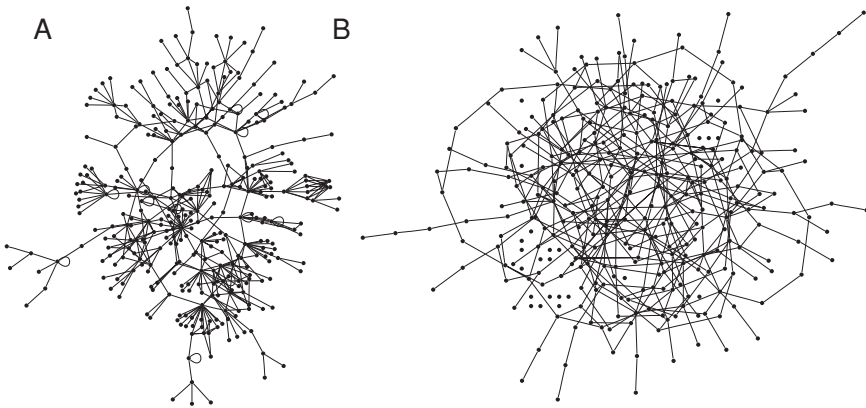


Figure 4.9. (A) The largest component of the known portion of the protein-protein interaction network of *Homo sapiens*. It is easy to see that this network is far from random. The majority of proteins interact just with a very limited number of other, such as 1 or 2, whereas a few proteins interact with tens of others. (B) A random graph with the same number of nodes and links that the network A, for comparison.

Whereas the random network has a more or less homogeneous distribution, the proteome network shows a huge number of proteins with a few connections, and at the same time, certain proteins have a very big number of connections, the so-called hubs. Other features can be distinguished right away, such as for example, the existence of groups of nodes that connect to two hubs at once.

To quantify the mentioned heterogeneity, the distribution of degree is shown, for the bigger proteome network of *Saccharomyces cerevisiae*, in Figure 4.10. Although many functions would qualify as degree distributions with a big number of low degree nodes and a small number of very high degree ones, the precise functional form of the distribution is a strict power-law of the form $P(k) \sim k^{-\gamma}$ with exponent between 2.1 and 2.5 (for different measuring methods and organisms, as shown in Figure 4.10). The networks in this case are much bigger, and hence, they have a broader distribution. The message is, then, clear enough: since the two distributions are hardly comparable, random networks seem to be poor models of real protein interaction networks. This is actually a more general result, which has been discovered in networks of very diverse fields [16,17].

To support the deviation from a purely random graph there is the fact that the proteome network is a small-world, since the comparison of its

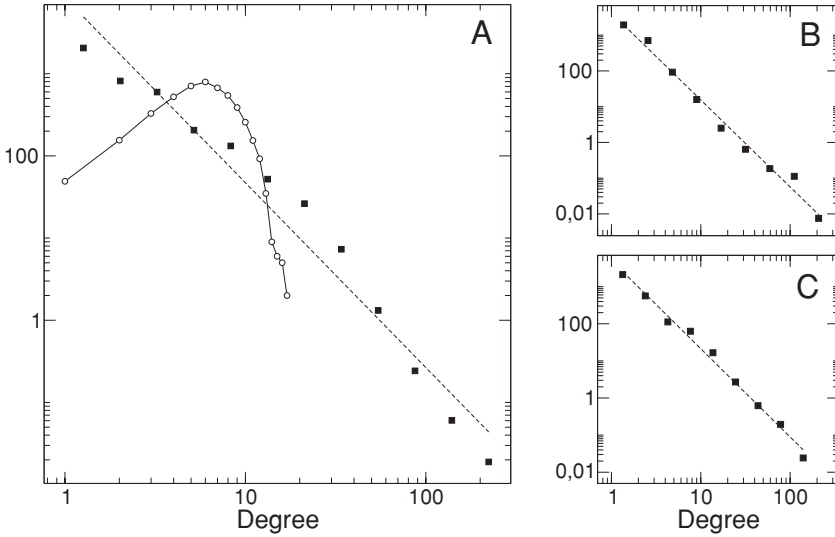


Figure 4.10. A. In black squares, degree distribution of the proteome graph of *Saccharomyces cerevisiae* taken from the DIP database and in white circles a random graph with the same number of vertices and edges. Note that axes are logarithmic, revealing that the proteome graph has a power-law degree distribution, that is $P(k) \sim k^{-\gamma}$. B *S. cerevisiae* from [18] ($N = 3280$, $L = 4549$ and $\gamma = 2.43 \pm 0.10$). C *C. elegans* taken from [19] ($N = 3228$, $L = 5625$ and $\gamma = 2.37 \pm 0.08$). To properly average the power laws, the bins were taken logarithmically distributed.

properties with those of a random graph of its size, that is,

$$\begin{aligned} C_{prot} &= 0.142 & L_{prot} &= 4.218 \\ C_{rand} &= 0.00139 & L_{rand} &= 4.515 \end{aligned}$$

in fact satisfy $C_{prot} \gg C_{rand}$ and $L_{prot} \approx L_{rand}$ (for this calculations the DIP dataset was used).

4.1. Assortativeness and correlations

The other mentioned feature seen in the graph of Figure 4.9A concerns the correlation between the degrees of vertices. Given two connected vertices in a network, we can simply ask what the correlation between their degrees is. In other words, assuming a certain degree distribution p_k to dictate the degrees of the vertices in an otherwise random network, we would expect the joint probability distribution of the remaining degrees of

vertices e_{jk} to be $q_j q_k$, the simple product of remaining degrees. If that is not the value found, we can measure the difference by averaging through all the edges in the graph, that is, $\langle jk \rangle - \langle j \rangle \langle k \rangle = \sum_{jk} jk(e_{jk} - q_j q_k)$. This sum measures the assortativeness of a network [20]. If it is positive, high degree nodes tend to connect to other high degree nodes, whereas a negative value describes the opposite, high degree nodes will preferentially be neighbors of low degree ones.

To properly compare different networks, it is convenient to normalize the sum by the maximum value attainable, so that the result is in the range $(-1, 1)$. The maximum value, in fact, is present when the joint distribution e_{jk} is simply $q_k \delta_{jk}$, that is, when nodes of a given degree just connect to nodes of the same degree. It is easy to see that the result is $\sum_k k^2 q_k - [\sum_k k q_k]^2$ which is the variance of the remaining degree distribution, σ_q^2 . Hence, the normalized correlation function is

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k). \quad (4.1)$$

When calculating this value for an actual network with known values, we can use the following formula [20],

$$r = \frac{L^{-1} \sum_i j_i k_i - [L^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{L^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [L^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2} \quad (4.2)$$

where j_i, k_i are the degrees of the vertices at the ends of the i th edge, with $i = 1, \dots, L$. When we apply this measure to protein interaction networks, the results are clear, protein networks are disassortative, that is, hubs connect with high correlation to low degree nodes. The results for the three networks in Figure 4.10 are, respectively, -0.1302 , -0.1361 and -0.1397 .

4.2. Correlation profiles

However, the assortativeness of a network gives just a global description of the correlations in connectivity. If we want more detail in this type of analysis, we have to turn to other techniques. One more time, measurements will involve the comparison of the correlations of a given graph with its randomized counterpart [21,22]. Since we are comparing degree correlations, though, it is important in this case to create a random graph that has the same degree distribution.

This is achieved through a rewiring process, in which the degree distribution is preserved. If two edges are chosen at random that do not have vertices in common, the simple exchange of their starting vertices will give a graph with the same degree distribution but otherwise random. Iterating this process as many times as twice the number of edges yields a reasonably randomized graph, with preserved degree distribution. If we call $P(k_0, k_1)$ the probability of finding an edge connecting two nodes with degree k_0 and k_1 , and $P_r(k_0, k_1)$ the random equivalent, we can measure two things, that is,

$$R(k_0, k_1) = \frac{P(k_0, k_1)}{P_r(k_0, k_1)}, \quad (4.3)$$

whose deviation manifest the correlations, and

$$Z(k_0, k_1) = \frac{P(k_0, k_1) - P_r(k_0, k_1)}{\sigma_r(k_0, k_1)}, \quad (4.4)$$

quantifying the statistical significance of $R(k_0, k_1)$, or Z-score. The value $\sigma_r(k_0, k_1)$ is the standard deviation of $P_r(k_0, k_1)$ in an ensemble of randomized networks.

The results of this process are shown in Figure 4.11. The image represents a log-binned matrix in which black indicates a higher value, as the side bar shows. This profile reveals, with some more precision, the disassortative nature of protein interaction networks.

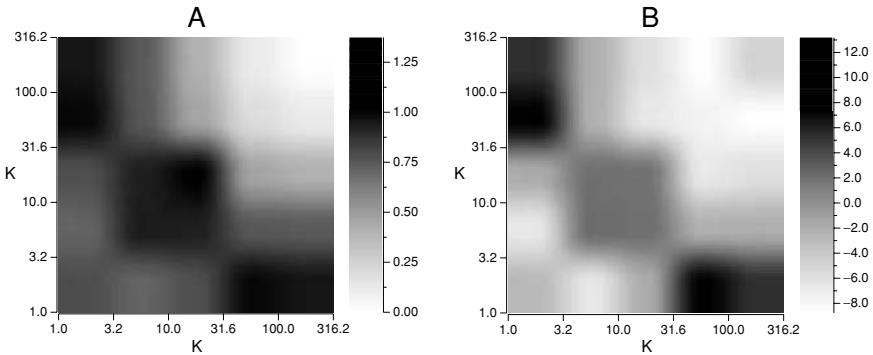


Figure 4.11. (A) Correlation profile of the protein interaction network, with the dataset taken from [18]. The value $R(k_0, k_1)$ is shown. (B) Same as A, but showing the Z-score, $Z(k_0, k_1)$.

To finish our analysis of the proteome network, we can see its modularity, which is something not apparent in the graph of Figure 4.9A, as we mentioned in the introduction. In the last section we saw that hierarchical clustering can give a clear picture of modularity lacking any good measure of it. A network such as the one shown in Figure 4.9A has a modularity

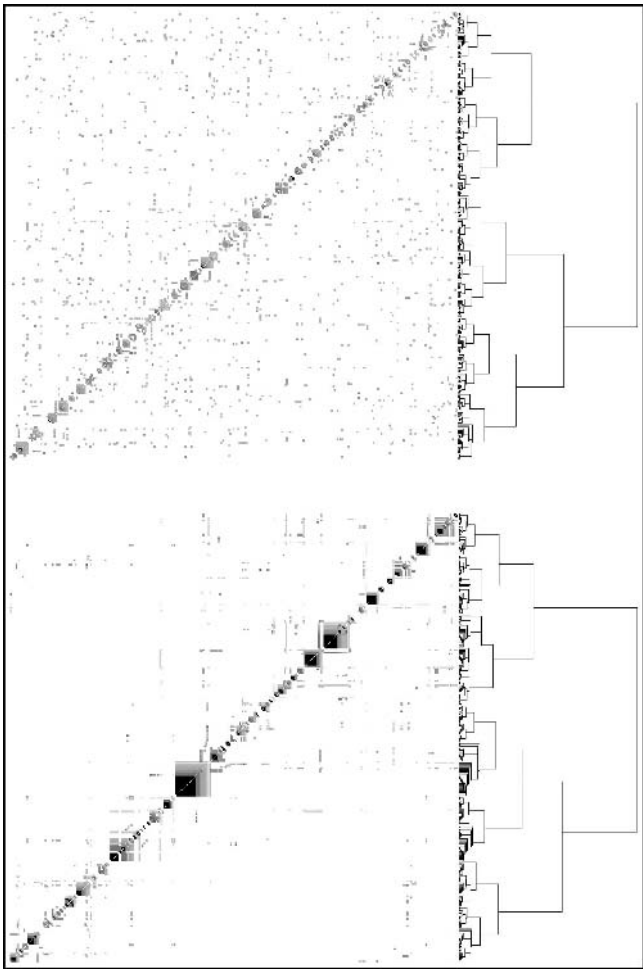


Figure 4.12. Bottom. Modular appearance of the network of Figure 9A. Top. The result of applying the same algorithm to a random network of about the same size.

that is revealed by the hierarchical clustering shown in Figure 4.12 (bottom). The random network of Figure 4.9B is also displayed for comparison (top).

4.3. Proteome model

Since we now know the structure of the proteome network, we can ask ourselves how the proteome network emerged. In principle, we can try to find simple mechanisms that we know alter the relationship between proteins in the genome and try to interpret them in terms of the rewiring process that affects the protein interaction network. A simple mechanism does, in fact, exist, and it gives a very plausible answer to the question of the origin of the proteome network.

The genome grows mainly due to gene duplication. From time to time, a gene is duplicated in the cell replication process that gives rise to a redundant copy of a gene. After a while, mutations accumulate that make the two copies diverge pushing their sequences apart, and as well the functionality of the proteins they code for. The result is that most genes currently present in the genome can be traced back to ancient duplications of other genes. Since proteins are the product of genes, we can model gene duplication in the way Figure 4.13 shows [23].

We start from a set m_0 of connected nodes, and at each time step we perform the following operations

- One node of the graph is selected at random and duplicated
- The links emanating from the newly generated node are removed with probability δ .
- New links (not previously present after the duplication step) are created between the new node and all any other node with probability α . Although available data indicate that new interactions are likely to be formed preferentially towards proteins with high degree here we do not consider this constraint.

Step (i) implements gene duplication, in which both the original and the replicated proteins retain the same structural properties and, consequently, the same set of interactions. The rewiring steps (ii) and (iii) implement the possible mutations of the replicated gene, which translate into the deletion and addition of interactions with different proteins, respectively. The process is repeated until N proteins have been obtained.

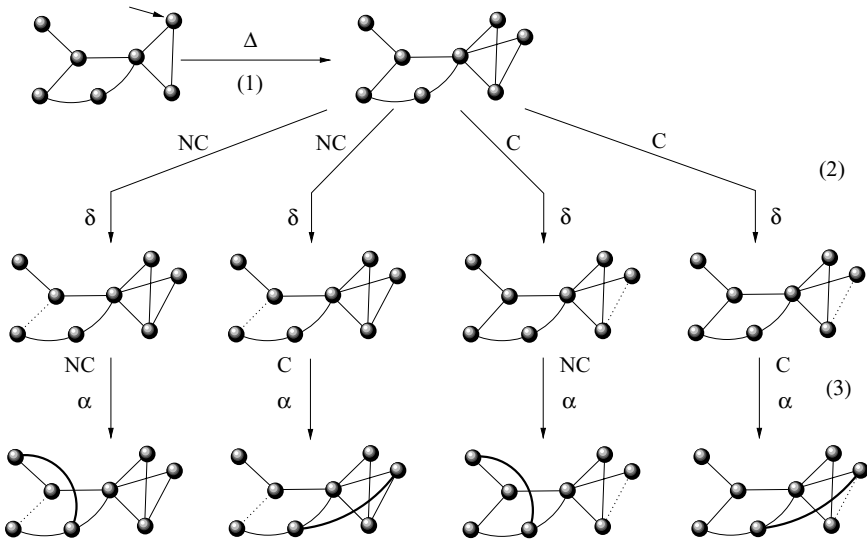


Figure 4.13. Rules of proteome growth in the four possible scenarios. First, (1) duplication occurs after randomly selecting a node (small arrow). Then (2) deletion of connections occurs with probability δ . This event can be correlated (C) when the deleted links are connected to the newly generated node or uncorrelated (NC), when all links are considered for deletion. Finally (3) new connections are generated with probability α , again in a correlated or uncorrelated way. The time scales at which different events occur are known to be very different: duplication takes place at a much slower rate, whereas rewiring is much faster. Additionally, the specific rates at which each event occur might involve preferential attachment to proteins of higher connectivity. All these variants can be included.

Another model uses very similar rules [24], but introduces some relevant differences. Duplication is also followed by two probabilistic rules which operate independently. The first (ii) is link deletion. For each of the nodes p_j linked to the two p_i and its duplicate p'_i , we choose randomly one of the two links $\varepsilon_{ji}, \varepsilon_{j'i'}$ and remove it with probability δ . Additionally, a new interaction connecting the two proteins (the parent and the duplicated) is introduced with probability ρ . The last rule will naturally increase the number of triangles in the system and thus provide a source of high clustering.

The rewiring process seems to be more appropriately defined, since the removal of one of the alternative links allows “conserving” the function that was somehow present before the duplication event. In Solé’s model, the whole set of links of the duplicated gene are preserved and loss of

connections affects only the new copy. By using Vázquez's approach, more flexibility is allowed and the interaction map is more likely to remain connected. As defined, it is important to note that duplicates will diverge only to some extent: if duplication occurs in a gene with degree k_i , only δk_i will be removed on average.

The two models collapse into a single mean field description where the average connectivity follows the dynamics

$$\frac{dK_n}{dn} = \frac{1}{n} (K_n + \phi_\alpha(n, K_n) - 2\delta K_n), \quad (4.5)$$

where $\phi = 2\alpha(n - K_n)$ in Solé's model and $\phi = 2\alpha(n - K_n) = \rho$ in Vázquez's model. Actually, in previous work [23] it is shown that in order to have convergence in the system towards a scale-free stationary distribution we need a very small rate of link addition (which is consistent with observations). If we assume that $\alpha \sim O(1/n)$ then a single link is added on average each step and thus the two models are identical in the low-addition limit. Specifically, if the graph is sparse, we have $\alpha(n - K_n) \approx \rho$, which results in a dynamical equation

$$\frac{dK_n}{dn} + \frac{2\delta - 1}{n} K_n = \frac{2\rho}{n} \quad (4.6)$$

which has an associated general solution

$$K_n = e^{-\eta(n)} \left(2\rho \int \frac{e^{-\eta(n)}}{n} dn + C \right) \quad (4.7)$$

where $\eta(n) = \int (2\delta - 1) dn/n = (2\delta - 1) \ln n$.

This gives

$$K_n = \frac{2\rho}{2\delta - 1} + \left(K_0 - \frac{2\rho}{2\delta - 1} \right) n^{-(2\delta - 1)} \quad (4.8)$$

if $\delta > \delta_c = 1/2$, the previous system converges to a graph with a finite average degree

$$K_\infty = \lim_{n \rightarrow \infty} K_n = \frac{2\rho}{2\delta - 1} \quad (4.9)$$

Otherwise, the average connectivity will be $K_\infty \rightarrow \infty$. The critical removal rate $\delta_c = 1/2$ thus defines a phase transition separating a phase with a highly-connected system ($\delta < \delta_c = 1/2$) from a sparse phase ($\delta > \delta_c$) where a finite number of links will be observed. At this phase, the network

becomes fragmented into many components. It is interesting to note that, under the present conditions, the long-term behavior of the average connectivity does not depend on the rate of link addition. What is really important is that the rate of link addition and link removal are similar, so that $\langle k \rangle$ can reach a stationary value. Moreover, it can be shown that although no explicit preferential attachment is included here, the multiplicative nature of the process (in which proteins having more links are more likely to have them copied) actually leads to an effective preferential attachment [25].

We can test this prediction by studying the behavior of the model under different rates of link deletion. In order to measure the impact of this rate on network's architecture, we use two different, but closely related measures: (1) the normalized largest component size S and (2) the average, normalized component size $\langle s \rangle$. If $C(\Omega) = \{\Omega_1, \Omega_2, \dots, \Omega_c\}$ is the set of connected components (subgraphs) of the proteome map, so that

$$\Omega = \bigcup_{i=1}^{\infty} \Omega_i \quad (4.10)$$

and $n_i = |\Omega_i|$ indicates their size (with $\sum_i n_i = N$), we define

$$S = \frac{1}{N} \max\{n_i\} \quad (4.11)$$

$$\langle s \rangle = \frac{1}{N} \left(\frac{1}{c} \sum_{i=1}^c n_i \right) \quad (4.12)$$

In Figure 4.14 we display the two measures against δ for a $N = 10^3$ protein network. Close to δ_c we can appreciate a clear change. The two phases are clearly identified, with the connected one showing $S \approx 1, \langle s \rangle \approx 1$ and the fragmented phase showing $S \approx 1/N, \langle s \rangle \approx 1/N$. In Figure 4.14 (left) we can see that S decreases slowly close to δ_c , where only about half of the nodes remain connected within the largest component. The sharpness of the transition becomes much more obvious in Figure 4.14 (right). Here we clearly appreciate the impact of rewiring on network's structure, indicating that a large fraction of the overall network structure is formed by small, isolated components. In Figure 4.15 we can see some examples of the graphs generated (largest components) obtained at different rates of deletion.

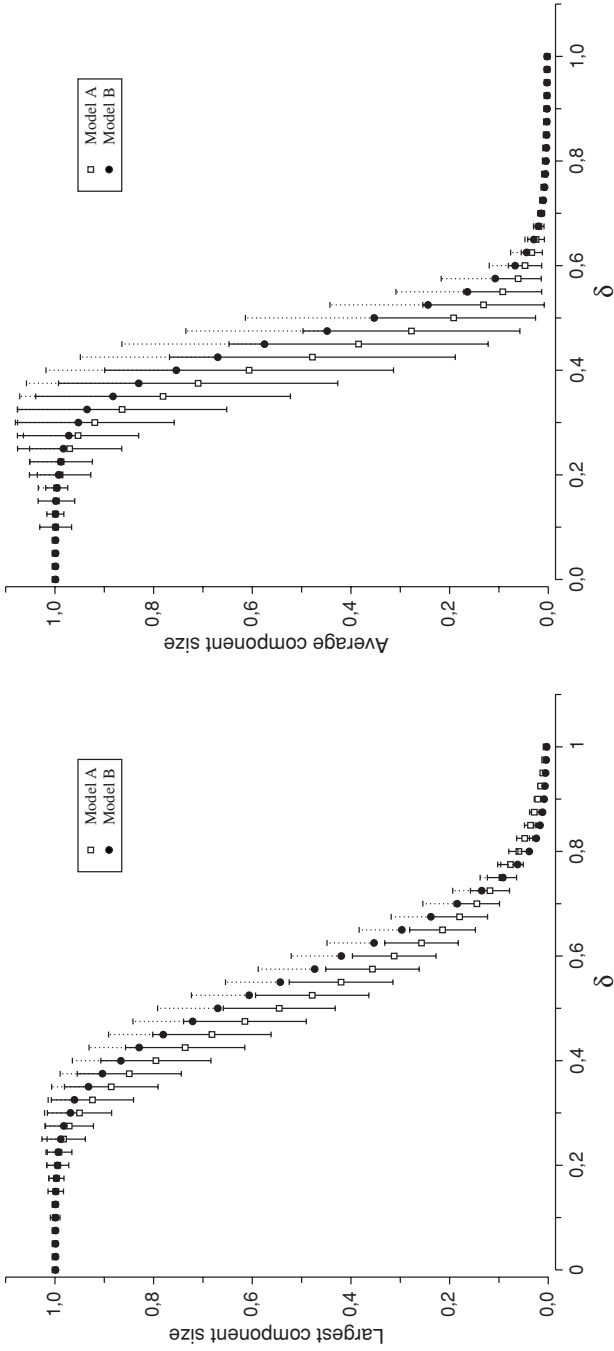


Figure 4.14. Phase transition in the genome growth models. Here $N = 10^3$ and averages have been performed over $R = 10^3$ replicas. Here the size of the largest component and the average size are shown against the rate of link removal δ . The predicted phase transition occurs at $\delta_c \approx 0.5$. Due to the finite (small) size of our networks, the transition appears to be less sharp than expected.

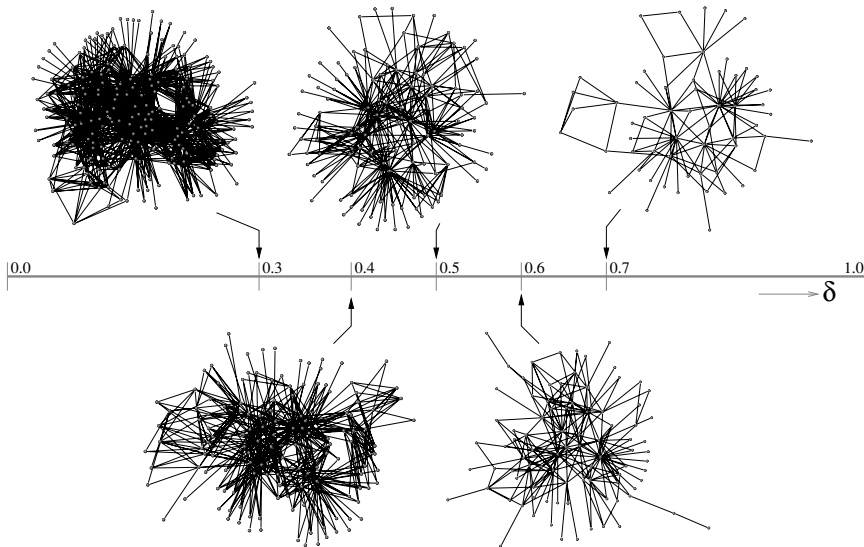


Figure 4.15. The architecture of the proteome map, as generated by the simple model for different values of the deletion rate δ . As predicted by the mathematical model, two well-defined phases are present. For the first, when $\delta < \delta_c = 0.5$, the protein map is highly connected and most elements have links to others. Conversely, for $\delta > \delta_c$, the graph is fragmented into many components and many components have no links or belong to small isolated subnets. Close to the transition domain, we have a sparse graph with the statistical features displayed by the real proteome map. Such graph displays modular organization, in spite of a complete lack of functionality in the definition of the model rules.

5. Gene Networks

The proteome network of last section demonstrates the importance of models that try to capture essential ingredients in network evolution. We have seen the networks implicit in protein structures, the networks that these molecules form when interacting within the cell as numerous complexes, but we are still lacking one important ingredient, and that is function. Although illuminating, previous analyses dealt only with topological properties, which can describe cells only partially. If we really want to understand cellular functioning, we must turn towards its function, and in particular, to regulatory networks.

It is known since long ago that genes interact with one another. This interaction is due to the fact that some proteins (transcription factors) can

bind to DNA and alter transcription of other proteins, modulating their concentrations. This regulation allows us to model a genome as a network: each gene is mapped to a node, and regulatory interactions are mapped to the edges between them. Since the existence of regulatory interaction from gene a to b doesn't necessarily imply also that b regulates a , we are in fact considering a directed network.

To be able to model this kind of networks, we can simplify things by considering a discretized version of a genome: although gene expression is known to be continuous, we can disregard this and consider genes Boolean. This approximation was pioneered by Kauffman more that 30 years ago [26]. This type of analysis, although to some purposes too simple, has been successful in models of genetic circuits [27,28]. In this kind of modeling, a gene is just like a switch, which can be turned on or off, that is, expressed or not expressed. The nodes in the network are then represented by a set of Boolean variables $\{\sigma_1, \sigma_2, \dots, \sigma_N\}$, which are, in fact, functions of discrete time. To determine the value of each gene in the next time step, we will use the values of the inputs of this gene, i.e., the values of the genes that regulate it, and to combine the diverse values of the inputs of a gene, a general function f_i is assumed to operate. In general, then, the dynamics of gene σ_i is

$$\sigma_i(t + 1) = f_i(\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_k}), \quad (5.1)$$

where k is the number of regulatory inputs of σ_i . In the general case, and without applying any explicit knowledge about the connectivity of real gene networks or the combinations of values performed in real genes, we can assume those to be random. First, the presence of an interaction between gene i and j will depend on a certain probability, which is the definition of a random graph, yielding a Poisson distribution for the number $\langle k \rangle$ of incoming links. Second, to make functions random, for every combination of values in the inputs, the value of the output will be 1 with probability p and 0 with probability $1 - p$. This leaves a control over the bias in the function, but leaves it otherwise random. Summarizing, we have a discrete dynamics over a random graph with random functions of Boolean values, a Random Boolean Network (RBN).

Since Boolean networks are deterministic systems, whenever a network reaches a state which has already been visited in a previous time-step, it will enter in a cyclic trajectory. Due to this fact, the whole state space with $\Omega = 2^N$ configurations is nicely partitioned into these cycles and their

basins of attraction, i.e., the sets of states that lead to them. Basically, this defines a different graph in which nodes are the global states of the network, and the links represent the transitions between states due to the dynamics of the Boolean network. The basins are, in fact, simply the components of this graph. Since the dynamics gets trapped into the cycles of each basin once they are reached, basins can be also described in terms of the length of the cycles and also the length of the transition branches leading to them.

Kauffman identified these cycles with the cell types of an organism with a genome modeled by the Boolean network [29], since they represent disjoint possible stable states in the dynamics of a single network (given that all cells in an organism have the same genome). Varying the connectivity $K = \langle k \rangle$, he found profound differences in numbers, lengths and transition times of the emerging cycles. For high connectivity, $K \geq 5$, cycles and transients are very long, and there is a rather small number of basins. Also, perturbing a gene in the network typically leads to another attractor. For low connectivity, $K = 1$, cycles are usually fixed points, and transients are short, with again a high sensibility to perturbations. But for $K = 2$, an unexpected order appears, what Kauffman called “order for free”: the network has a small number of cycles and they are very robust against perturbations.

All these interesting properties about the cycles in the network can be explained by the existence of a phase transition [30], that is, a sudden change in behavior for a smooth change in parameters. The only parameters in this model are $\langle k \rangle$ and p , which govern connectivity and the function bias, respectively. This allows us to draw a phase diagram, shown in Figure 4.16 (left), which clearly separates the three distinct phases: ordered, chaotic, and critical. The critical phase is actually the interphase between the ordered and the chaotic one, whose threshold is defined by the equation

$$Kc = \frac{1}{2p(1-p)}. \quad (5.2)$$

The names come from the observation of the dynamics in each case, which is also depicted in Figure 4.16 (right). If we take a constant value of $p = 1/3$ and just move the parameter K (just as Kauffman did, although with $p = 1/2$), we find that for low values of K , below 2.25, the dynamics is frozen, and the values of gene expression stop switching altogether (Figure 16A). This behavior resembles that of cycles. For higher values, above 2.25, we find that the dynamics is completely noisy and has no coherence (Figure. 4.16C), which also would explain the cycle behavior. For exactly

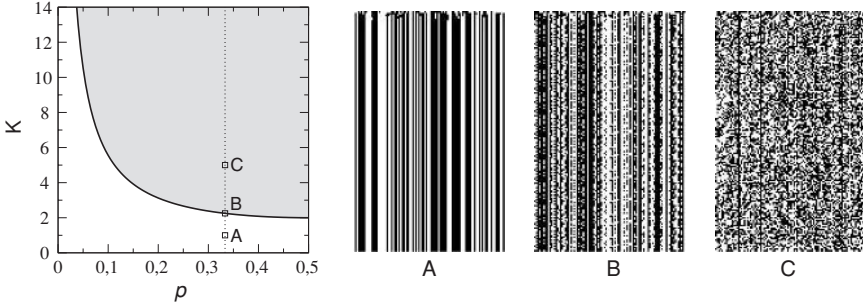


Figure 4.16. Left. Phase diagram of a random Boolean network of parameters K and p . The white area is the ordered regime, the grey area is the chaotic regime, and the critical phase is the black line in between. Right. Temporal dynamics of the genes in the network at the points A, B and C along the line $p = 1/3$ specified in the phase diagram. (A) Ordered, (B) Critical, and (C) Chaotic.

the value $K_c = 2.25$, we can see that the dynamics is ordered and still not totally frozen, but with some complex periodicity (Figure 4.16B).

To approach this phase transition mathematically, we can perform a perturbative analysis, that is, we can ask what the consequence of flipping the state of a small fraction of genes in the network is. If these changes in value affect other genes that have the damaged genes as inputs, these initial “errors” will propagate through the network, spreading the damage. Let us, then, consider the difference $\delta(t)$ of two configurations of values $\Sigma(t) = \{\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)\}$ and $\Sigma'(t) = \{\sigma'_1(t), \sigma'_2(t), \dots, \sigma'_N(t)\}$, which is defined as

$$\delta(t) = \frac{1}{N} \sum_{i=1}^N |\sigma_i(t) - \sigma'_i(t)|. \quad (5.3)$$

If the temporal evolution of these two configurations is governed by the same Boolean network, with identical regulatory links and functions, the new configurations $\Sigma(t+1)$ and $\Sigma'(t+1)$ will define a new value $\delta(t+1)$, which will indicate the tendency to grow or shrink of the differences. A tendency to shrink or grow would be identified as the ordered phase or chaotic phase, respectively, with the critical, maintaining phase in between, in which the damage neither expands nor shrinks on average.

To see what is the tendency of $\delta(t)$, we can look at the change in value of the units with inputs affected by changes, that is, if

$$f_i(\sigma_{i_1}(t), \dots, \sigma_{i_k}(t)) = f_i(\sigma'_{i_1}(t), \dots, \sigma'_{i_k}(t)) \quad (5.4)$$

with the probability that $\sigma_k(t) \neq \sigma'_k(t)$ being $\delta(t)$. Since f_i values are uncorrelated, any number of changes in the inputs of a gene will give a different, random value, and therefore the probability that the output will change is $2p(1 - p)$ (the probability that the $f(\Sigma(t))$ is different than $f(\Sigma'(t))$). If we use this for the dynamics of $\delta(t)$ we get

$$\begin{aligned} \delta(t + 1) &= \sum_{k=1}^{\infty} 2p(1 - p) [1 - [1 - \delta(t)]^k] P_{in}(k) \\ &= 2p(1 - p) \left[1 - \sum_{k=1}^{\infty} [1 - \delta(t)]^k P_{in}(k) \right] \end{aligned} \quad (5.5)$$

which is the same as saying that

$$\delta(t + 1) = M(\delta(t)) \quad (5.6)$$

with the mapping $M(x)$ being

$$M(x) \equiv 2p(1 - p) \left[1 - \sum_{k=1}^{\infty} (1 - x)^k P_{in}(k) \right]. \quad (5.7)$$

In the limit $t \rightarrow \infty$, the difference $\delta(t)$ will tend to the fixed point of equation $x = M(x)$. Even if $x = 0$ is always a fixed point of Eq. (5.6), its stability depends on $P_{in}(k)$. Since M is a monotonically increasing function of x , and also that $M(0) = 0$ and $M(1) = 2p(1 - p)$, Eq. (5.6) will have a fixed point $x^* \neq 0$ only if $\lim_{x \rightarrow 0^+} M'(x) > 1$. In fact, this condition is the threshold for the ordered to chaotic transition. Figure 4.17 depicts the function $M(x)$ for different values of p (the same as in Figure 4.16), showing how $K_c = 2.25$ is tangent to $f(x) = x$ at 0.

From Eq. (5.7) it follows that

$$\lim_{x \rightarrow 0} \frac{dM(x)}{dx} = 2p(1 - p) \sum_{k=1}^{\infty} k P_{in}(k) = 1 \quad (5.8)$$

which is exactly the same as Eq. (5.2) above. As we can see, the critical point for the dynamical transition in Boolean networks depends only on the average value of the distribution $P_{in}(k)$, and not on its exact shape. This result is useful, therefore, for any degree distribution, which is important since genetic regulatory networks are not really random.

Available data has confirmed that genetic regulatory networks are indeed scale-free [32]. Figure 4.18 shows a small graph corresponding to

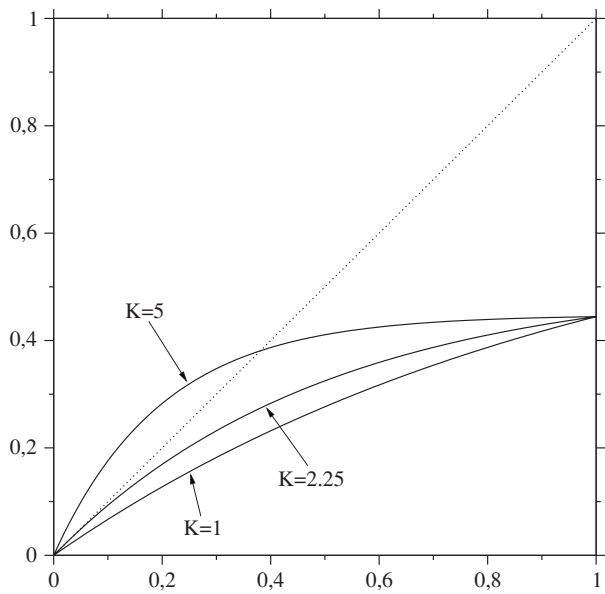


Figure 4.17. Graphical representation of $M(x)$ for different values of K . The values are the same as in Figure 16, showing chaotic phase $K = 5$, critical phase $K_c = 2.25$, and ordered phase $K = 1$. The circles show the fixed points of equation $x = M(x)$.

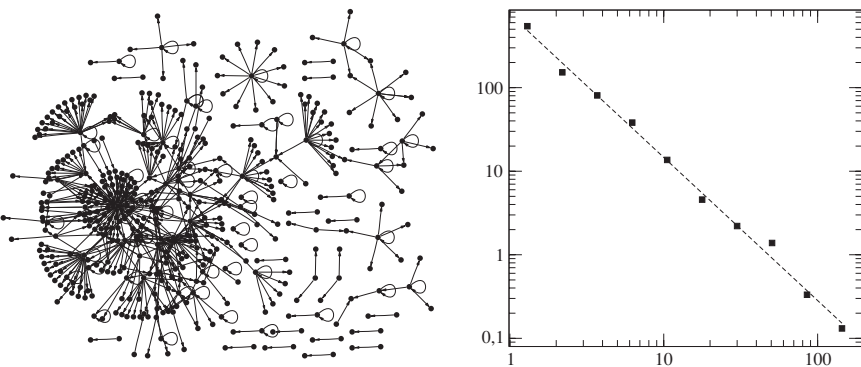


Figure 4.18. Left. Directed graph corresponding to a portion of the genetic regulatory network of *E. coli*. Right. Degree distribution of the genetic regulatory network of yeast [31].

the regulatory network of *E. coli* (left) and the degree distribution for a the bigger network of *S. cerevisiae* (right). This new finding has stimulated research in Boolean networks which have a power-law degree distribution [33,34], further extending the results on random Boolean networks.

In a scale-free Boolean network, as mentioned, the condition of Eq. (5.2) is still valid, and for a degree distribution $P(k) = Z(\gamma)k^{-\gamma}$, the average degree is determined by the expression

$$\langle k \rangle = \sum_{k=0}^{\infty} k P(k) = \frac{Z(\gamma - 1)}{Z(\gamma)}, \quad (5.9)$$

where $Z(\gamma)$ is the Riemann Zeta function, defined as $Z(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$. Substituting this value in Eq. (5.2) yields a transcendental equation involving γ . Using γ as a descriptor of a power law distribution makes sense, since the variance in this kind of distributions is not bounded for $\gamma < 3$, and hence the average degree is not a meaningful measure to characterize $P(k)$. The corresponding phase diagram using γ (instead of K) and p can be seen in Figure 4.19.

In the light of these results, it is conjectured that the extremely heterogeneous degree distribution of a scale-free network can provide a genome

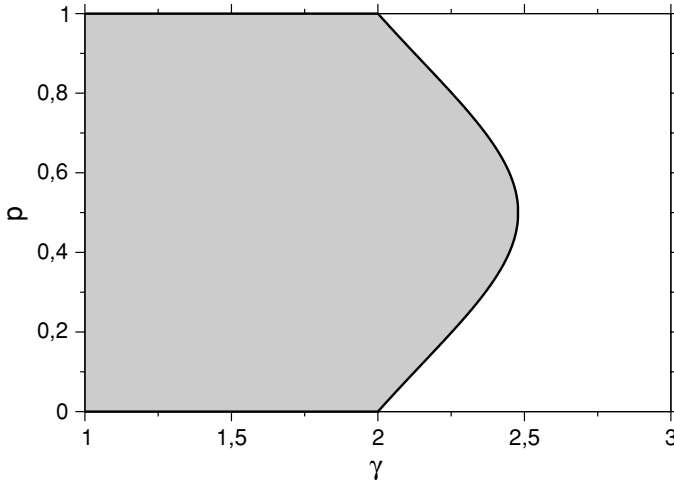


Figure 4.19. Phase diagram of the dynamical phase transition of a Boolean network with $P(k) = Z(\gamma)k^{-\gamma}$ as a function of γ and p , to be compared with diagram in Figure 16 (left). The grey area corresponds to the chaotic phase, and the white region to the ordered phase, with the black line being the critical phase.

with two important properties at once [33]. First, the robustness associated with biological systems that the abundant ordered regime can provide. Second, due to the sensitivity of the network to changes of function or connectivity in high-degree nodes, the evolvability of the network is guaranteed. This stands in contrast to the insensitivity of a more homogeneous network, in which robustness necessarily makes adaptation very difficult.

6. Overview

Exploring the static, graph-level structure of a complex system is always limited by the many details that are missed and also by the lack of functionality. However, networks reveal a pattern that results from the process of evolution and this imposes several constraints on the possible explanatory mechanisms. Nevertheless, even the architecture of a network can reveal key features of the underlying functional organization. The modules found in protein contact maps were associated to well defined domains, and the details revealed by the hierarchical clustering algorithm actually suggest that we are not facing a simple modular pattern, but instead a hierarchical, nested architecture.

What can be learned about network origins from network structure? A standard view of evolution would suggest that modules might have resulted from a process of selection in which the advantages introduced by specialized compartments would be favored by selection. But the models of proteome evolution suggest instead that given that a critical balance in connectivity is present, the patterns observed all emerge, including modularity.

When using a simple model of functionality in networks, such as Boolean models, we again face the underlying constraints imposed by the architecture. If we want to make a regulatory network organized, we have to choose from a reduced set of values for the sensitivity of nodes and the connectivity. Theory seems to illuminate very well the question of what are the possible ways in which a regulatory network could be organized. As in the case of protein networks, not all options are available.

Using networks as a theoretical framework seems to be relevant for a number of reasons. The first is that they provide well-defined quantitative properties to be reproduced by dynamical models of network evolution. Moreover, it is becoming clear that many network properties might be unique in terms of the universe of possible patterns of interaction among

units in a complex network: once some minimal requirements such as a connected network with sparse connectivity are reached, other key features might emerge “for free”. In such a view of network complexity, the architecture would be to a large extent a blueprint of underlying laws of organization. Dynamics, selection and adaptation would be shaped by such architectural constraints.

Acknowledgements

The authors would like to thank the members of the complex systems research group for useful discussions. This work was supported by grant BFM2001-2154, EU PACE grant (Programmable Artificial Cell Evolution) FP6-002035, the Generalitat de Catalunya (PFD, 2001FI/00732), and The Santa Fe Institute.

References

1. R. V. Solé, R. Ferrer, J. M. Montoya and S. Valverde, *Complexity* **8**, 20-33 (2002).
2. B. Bollobás, *Random Graphs*, 2nd edition, Cambridge University Press, Cambridge, UK (2001).
3. M. E. J. Newman, Random Graphs as Models of Networks, in *Handbook of Graphs and Networks*, S Bornholdt and HG Schuster (eds.), John Wiley-VCH, Berlin (2002) pp. 147-169.
4. D. Watts and S. H. Strogatz, *Nature* **393**, 440-442 (1998).
5. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th edition, Garland Science, New York (2002).
6. A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, *PNAS* **92**, 1282-1286 (1995).
7. J. Espadaler, N. Fernández-Fuentes, A. Hermoso, E. Querol, F. X. Aviles, M. J. Sternberg, and B. Oliva, *Archdb, Nucl. Acids. Res.*, **32D**, 185-188 (2004).
8. L. W. Ancel and W. Fontana, , *J. Exp. Zool.* **288**, 242-283 (2000).
9. G. von Dassow, E. Meir, E. Munro and G. M. Odell, *Nature* **406**, 188-192 (2000).
10. G. von Dassow and E. Munro, *J. Exp. Zool.* **406**, 188-192 (1999).
11. R. V. Solé, I. Salazar-Ciudad, and S. A. Newman, *Trends Ecol. Evol.* **15**, 479-480 (2000).
12. G. P. Wagner, *Am. Zool.* **36**, 36-43 (1996).
13. G. P. Wagner and L. Altenberg, *Evolution* **50**, 967-976 (1996).
14. N. M. Stadler, P. F. Stadler, G. P. Wagner, and W. Fontana, *J. Theor. Biol.* **213**, 241-274 (2001).
15. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási, *Science* **297**, 1551-1555 (2002).
16. R. Albert and A. L. Barabási, *Rev. Mod. Phys.* **74**, 47-97 (2002).
17. S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks. From Biological Nets to Internet and WWW*, Oxford University Press, Oxford (2003).

18. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, *PNAS* **98**, 4569-4574 (2001).
19. S. Li, C. M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P. O. Vidalain, J. D. Han, A. Chesneau, T. Hao, D. S. Goldberg, N. Li, M. Martinez, J. F. Rual, P. Lamesch, L. Xu, M. Tewari, S. L. Wong, L. V. Zhang, G. F. Berriz, L. Jacotot, P. Vaglio, J. Reboul, T. Hirozane-Kishikawa, Q. Li, H. W. Gabel, A. Elewa, B. Baumgartner, D. J. Rose, H. Yu, S. Bosak, R. Sequerra, A. Fraser, S. E. Mango, W. M. Saxton, S. Strome, S. Van Den Heuvel, F. Piano, J. Vandenhaute, C. Sardet, M. Gerstein, L. Doucette-Stamm, K. C. Gunsalus, J. W. Harper, M. E. Cusick, F. P. Roth, D. E. Hill, and M. Vidal. *Science* **303**, 540-543 (2004).
20. M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
21. S. Maslov and K. Sneppen, *Science* **296**, 910-913 (2002).
22. M. Maslov, K. Sneppen and U. Alon, Correlation Profiles and Circuit Motifs in Complex Networks, S. Bornholdt and H. G. Schuster (eds.), *Handbook of Graphs and Networks*, John Wiley-VCH, Berlin (2002) pp. 168-198.
23. R. V. Solé, R. Pastor-Satorras, E. Smith and T. Kepler, *Adv. Complex Systems* **5**, 43-54 (2002).
24. A. Vázquez, *Phys. Rev. E* **67**, 056104 (2003).
25. A. Vázquez, A. Flammini, A. Maritan, and A. Vespigniani, *ComplexUs* **1**, 38-44 (2003).
26. S. A. Kauffman, *J. Theor. Biol.* **22**, 437-467 (1969).
27. R. Albert and H. G. Othmer, *J. Theor. Biol.* **223**, 1-18 (2003).
28. R. V. Solé, P. Fernández, and S. A. Kauffman, *IJDB* **47**, 685-693 (2003).
29. S. A. Kauffman, *The Origins of Order*, Oxford U. Press, New York (1993).
30. M. Aldana-González, S. Coppersmith, and L. P. Kadanoff, Boolean Dynamics with Random Couplings, in *Perspectives and Problems in Nonlinear Science, A Celebratory Volume in Honor of Lawrence Sirovich*, E. Kaplan, J. E. Marsden, and K. R. Sreenivasan (eds.), Applied Mathematical Sciences, Springer, New York (2002).
31. A. H. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz, R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Menard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A. M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone *Science* **303**, 808-813 (2004).
32. D. E. Featherstone and K. Broadie, *BioEssays* **24**, 267-274 (2002).
33. M. Aldana, *Physica D* **185**, 45-66 (2003).
34. M. Aldana and P. Cluzel, *PNAS* **100**, 8710-8714 (2003).

Chapter 5

QUANTITATIVE MEASURES OF NETWORK COMPLEXITY

Danail Bonchev and Gregory A. Buck

Center for the Study of Biological Complexity, Virginia Commonwealth University Richmond,
Virginia 23284-2030

1. Some History

The first attempts to evaluate quantitatively the complexity of a system have been related to complexity of cells, organisms, and humans. Fascinated by the complex nature of the living things, a group of young mathematical biologists applied in the 1950s the Shannon theory of communications [1] to assess the information content of the living matter [2-5]. The analysis made by Rashewsky [4] provided the first proof that life on earth cannot emerge as a random event, because the probability for such an event would be incredibly small. Two different approaches have been used in defining the information content. The first one proceeded from the elemental composition of the living matter (C, N, O, etc.) and is the predecessor of what is nowadays called *compositional complexity*. Rashewsky's *topological information* has been based on partitioning the atoms in a structure according to both their chemical nature and their equivalent topological neighborhoods. Mowshovitz [6] developed further these ideas to define complexity of graphs. Minoli[7] introduced his *combinatorial complexity* of graphs, proceeding from the count of the graph vertices, edges, and paths.

In parallel with these attempts, another definition of information content has been advanced by Kolmogorov [8]. His *algorithmic information* has been defined as the minimal length of the program that exhaustively describes a given system. This type of information measure has found a broad application in computer sciences. The relevance of algorithmic information in describing *structural* complexity, however, is low [9], which limited its application to chemistry, whereas in biology it has found some application in assessing the genome complexity.

Shannon's information has been widely applied in chemistry in the form of information indices, characterizing different aspects of chemical structure [10-16]. These structural descriptors have been commonly used for quantitative structure-property and structure-activity relationships (QSPR and QSAR). However, only few of them have satisfied the requirements for a complexity measure [17]. Bertz introduced in 1981 his molecular complexity index applying Shannon's equation to the distribution of the two-edge subgraphs in molecular graphs [18]. That was the starting point of a systematic search in chemical theory for relevant measures of molecular complexity, a search that shifted the focus from information theory to molecular topology and graph theory. A series of requirements have been formulated for a structural descriptor to be a complexity measure [19-21], along with hierarchical concepts of molecular complexity [22,23]. A number of high quality measures of topological complexity have been devised during the last 7-8 years [24-31]. Complexity of chemical reaction networks has also been addressed making use of the spanning subgraphs of these cyclic graphs [32-35].

In the meantime, in the middle of 1980s, complexity theory emerged as a new integrative branch of science. The emphasis in the new theory was put on the complex dynamic systems, systems characterized by non-linear dynamics and emergent events. The quantitative aspects of the theory, related to random graphs, did not bring exciting results. The situation changed radically only when it was realized that any dynamic evolutionary system could be adequately presented by a network (a graph) that is non-random. Thus, complexity theory has found its universal language to describe systems as diverse as discrete space-time, the living cell, ecosystems, financial markets, World Wide Web, and social systems. This opened the door to the introduction of general methods for characterizing systems complexity, not only as information-based compositional complexity but, most essentially, as topological complexity of the network representing the system.

This chapter aims at elucidating the methods for quantitative assessments of networks complexity. It borrows from the rich arsenal of such methods developed during the last 25 years in chemical graph theory and chemical information theory. Being devised in a sophisticated way so as to distinguish the complexity of the multitude of molecules, these methods will be presented in a form adapted to the very large size of networks in

biology and ecology. New graph invariants having properties of complexity measures will also be presented. Examples of cellular and ecological networks will be analyzed with the methods presented.

2. Networks as Graphs

Networks are well characterized both quantitatively and as structural patterns or motifs by graph theory, which has at least 150 years of extensive development and application. Graph theory as a branch of discrete mathematics has been brought to life to solve specific problems from three different areas of science. Leonard Euler in 1788 constructed the first graph to solve the famous mathematical puzzle for the Königsberg bridges, a problem that is a predecessor of the transport and communication sets problems of our time. Rudolf Kircchoff in mid 19th century reinvented graphs and developed their theory to solve fundamental problems of electrical sets, a work of great value for the electronic networks of the 21st century, as well as for the complex chemical reaction networks. The third root of graph theory is in structural chemistry, which in the last part of 19th century was trying to determine the number of isomers, chemical compounds having the same atomic composition but different spatial structure.

The variety in the graph theoretical background produced a variety of non-standardized terminologies. In this chapter, we shall follow mainly the manner the terminology is used in chemical graph theory. Cellular networks are molecular networks, and we believe that the use of terms like “wirings” coming from electrical and computer engineering should be avoided in describing living things. This section introduces some basic graph theoretical notions and descriptors needed for the network topological and complexity analysis.

2.1. Basic notions in graph theory [36-38]

A network is defined by the set of V *vertices* (nodes, points), $\{V\} \equiv \{v_1, v_2, \dots, v_V\}$, and the set of E *edges* (links, lines), $\{E\} \equiv \{E_1, E_2, \dots, E_E\}$. The edge $\{ij\}$ is the line that emanates from vertex i and ends in vertex j . A *subgraph* is a graph obtained from the parent graph by deleting at least one edge or a vertex with its incident edges. A *loop* is an edge that begins

and ends in the same vertex. A *multigraph* is a graph in which some pairs of vertices are linked by more than one edge. *Simple graphs* are graphs having no multiple edges and loops. In a *complete graph*, K_V , any two vertices are connected by an edge. A *directed graph* is a graph having at least one directed edge. Directed edges are termed *arcs*. Graph without any directed edge is *undirected*. The graph is *connected* when there is a path between any pair of vertices in it; otherwise the graph is *disconnected*. A *path* in the graph is a sequence of adjacent edges without traversing any vertex twice. A *path graph*, P_V , is a graph containing only one path. A *star-graph*, S_V , is a graph containing one central vertex and $V-1$ branches of length one edge. A *walk* is an alternating sequence of vertices and edges, each of which could be traversed more than once. The *walk length* is the number of edges in it. A *cycle* is a path that starts from and ends in the same vertex. Graphs containing at least one cycle are called *cyclic graphs*. *Trees* are graphs containing no cycles. A *spanning tree* is a connected acyclic graph containing all the vertices of the graph. Graph *components* are connected subgraphs or vertices that are not connected to each other. Euler's theorem relates the number of vertices V , edges E , independent cycles C , and components K :

$$C = E - V + K \quad (2.1)$$

Figure 5.1 illustrates the notions introduced.

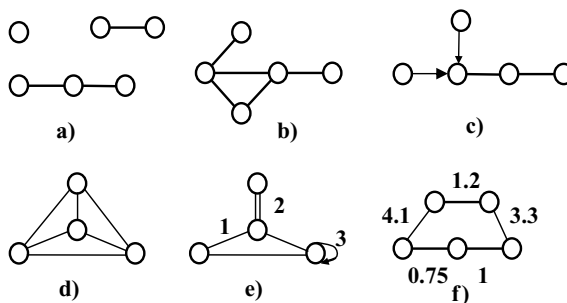


Figure 5.1. a) A disconnected graph with three components. b) A simple connected undirected graph. c) A directed graph. d) A complete graph with three cycles (the enveloping cycle is not counted, because it is not an independent cycle). e) A multigraph with a loop: 1, edge; 2, double edge; 3, loop. f) A weighted graph.

2.2. **Adjacency matrix and related graph descriptors**

Two vertices j and i are called *adjacent* when they are connected by an edge $\{i,j\}$. The adjacency relation is quantified by the term $a_{ij} = 1$, and the no adjacency one by $a_{ij} = 0$. The number of the nearest-neighbors of a vertex i is termed *vertex degree*, a_i . *Vertex degree distribution* is an ordered, usually descending set of vertex degrees, $\{V_{ord}\} \equiv \{v_{max}, \dots, v_{min}\}$. The sum of all vertex degrees in a graph defines its *total adjacency*, A . The matrix containing all adjacency relations in a graph G is called *adjacency matrix*, $A(G)$. The vertex degree of vertex i is calculated as the sum over all entries in the i^{th} row of adjacency matrix. Similarly, the total adjacency of graph G , $A(G)$, is calculated also as the sum over all matrix elements, a_{ij} :

$$a_i = \sum_{j=1}^V a_{ij}; \quad A(G) = \sum_{i=1}^V \sum_{j=1}^V a_{ij} = \sum_{i=1}^V a_i \tag{2.2a,b}$$

Undirected graphs (G) have adjacency matrices that are symmetrical with respect to their main diagonal, $a_{ij} = a_{ji}$. In directed graphs (DG), the symmetry of adjacency matrix is destroyed. Examples are shown in Fig. 5.2.

The vertex degrees of graph **2** shown below are actually *out-degrees*; they count the outgoing edges but not the incoming ones. Similarly, $A(2)$

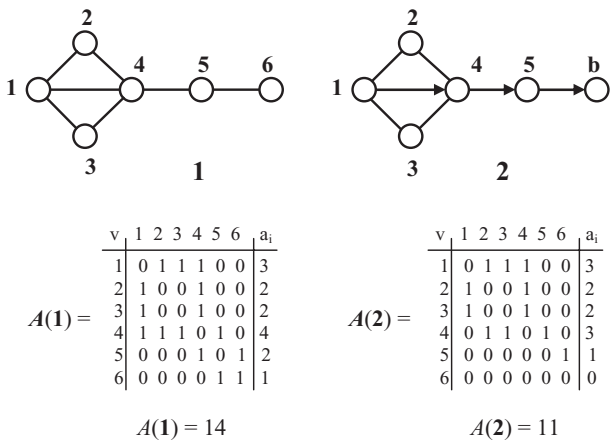


Figure 5.2. The undirected graph **1**, the directed graph **2**, their adjacency matrices $A(1)$ and $A(2)$, and total adjacencies $A(1)$ and $A(2)$, respectively.

out-degree one $\{3, 3, 2, 2, 1, 0\}$. The *in*-degrees are calculated by summing over each column, which for vertices 1 to 6 results in the set of *in*-degrees 2, 2, 2, 3, 1, 1, producing again $A_{in}(2) = 11$. One may generalize that the *in*- and *out*-adjacencies of the directed graph are equal and smaller than the adjacency of the parent undirected graph:

$$A_{out}(DG) + A_{in}(DG) = A(G) \quad (2.3)$$

The adjacency matrix of a graph provides also some generalized descriptors of network connectivity like the *average vertex degree* $\langle a_i \rangle$ and *connectedness* (or connectance), *Conn*:

$$\langle a_i \rangle = \frac{A}{V}; \quad Conn = \frac{A}{V^2} = \frac{2E}{V^2} \quad (2.4a,b)$$

For the undirected graph shown above, Eq. (2.4) produces $\langle a_i \rangle = 14/6 = 2.333$, and $Conn = 14/36 = 0.389$ (or 38.9%). The directed graph is less connected than the undirected graph with the same number of vertices and edges, as can be seen from the values obtained, $\langle a_i \rangle = 1.833$ and $Conn = 0.306$, respectively.

When dealing with undirected graphs, connectedness is frequently defined slightly differently as $Conn' = 2E / V(V-1)$. Here, $V(V-1)/2$ is the number of edges in the maximally connected graph (*complete graph*) having the same number of vertices. Connectedness is therefore a measure for the relative graph connectivity defined within the 0 to 1 range (or within the 0-100% range, after multiplying by 100). Formula (4b) defines graph connectedness in a more general manner, taking into account also the potential availability of non-zero diagonal adjacency matrix entries, $a_{ii} = 1$. The total number of matrix entries in this case is V^2 , not $2E / V(V-1)$. A non-zero diagonal element of adjacency matrix stands for a *loop*, which is an edge emanating from and ending in the same vertex. A loop represents self-interaction of the species described by the network nodes. Such are, for example, protein dimers in protein-protein networks, cannibalistic species in ecological food webs, and others.

2.3. Clustering coefficient and extended connectivity

The vertex degree a_i , which counts the nearest neighbors of a vertex i , is not the only *local* connectivity descriptor. More detailed information on the vertex neighborhood is contained in the *clustering coefficient*, c_i . It is

defined as the ratio of the number of edges E_i between the first neighbors of the vertex i , and the maximum number of edges, $E_i(\max) = a_i(a_i - 1)/2$, in the complete graph that can be formed by the nearest neighbors of this vertex:

$$c_i = \frac{2E_i}{a_i(a_i - 1)} \quad (2.5)$$

Applying Eq. (2.5) to the nondirected graph **1** shown in the foregoing, one obtains for the clustering coefficients the values $c_5 = c_6 = 0$, $c_4 = 1/3$, $c_1 = 2/3$, and $c_2 = c_3 = 1$. In the corresponding directed graph **2**, the clustering coefficient of vertex 4 goes down to zero.

More detailed description of graph connectivity takes into account the second and further neighborhoods. This can be done both locally and globally. The *second clustering coefficient* c_i' counts the edges between the second neighbors of vertex i , and again compares that count to the number of edges in the complete graph that could be formed by all second neighbors. Globally, the layers of second, third, etc., neighbors are taken into account in calculating the graph n^{th} -order extended *connectivity* [39], nEC . The calculation is performed by an iterative procedure, which at each step recalculates the vertex degree of each vertex as the sum of vertex degrees of its first neighbors, as obtained in the previous iteration:

$${}^nEC = \sum_{i=1}^V {}^na_i = \sum_{i=1}^V \sum_{j \text{ adj } i} {}^{n-1}a_j \quad (2.6)$$

One may thus form a vector of the extended connectivities of increasing order, $\{EC\} \equiv \{{}^0EC, {}^1EC, {}^2EC, \dots\}$, the zero-order term in which is the total graph adjacency, defined by Eq. (2.2b). Illustration of the iterative calculation of the first several kEC – terms of graph **1** is shown in Figure 5.3.

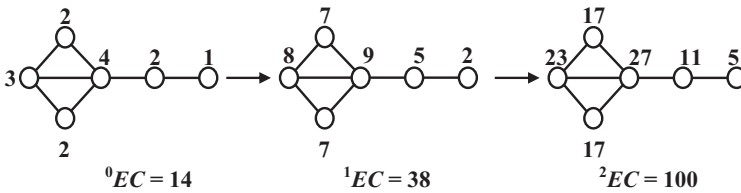


Figure 5.3. Iterative calculation of the first- and second-order extended connectivity of graph **1** (The null-order is identical to the total adjacency of the graph).

2.4. Graph distances

In Section 2.1, a *path* in the graph was defined as a sequence of adjacent edges between two vertices without traversing any intermediate vertex twice. The *distance* d_{ij} between vertices i and j is the shortest path between them. The *distance matrix* $D(G)$ of graph G is a square $V \times V$ matrix, which for undirected graphs is symmetrical with respect to the main diagonal. The sum over the matrix row entries is termed *vertex distance degree* of simply *vertex distance*, d_i . The sum over all distance matrix entries is called *graph distance*, D :

$$d_i = \sum_{j=1}^V d_{ij}; \quad D(G) = \sum_{i=1}^V \sum_{j=1}^V d_{ij} = \sum_{i=1}^V d_i \quad (2.7a,b)$$

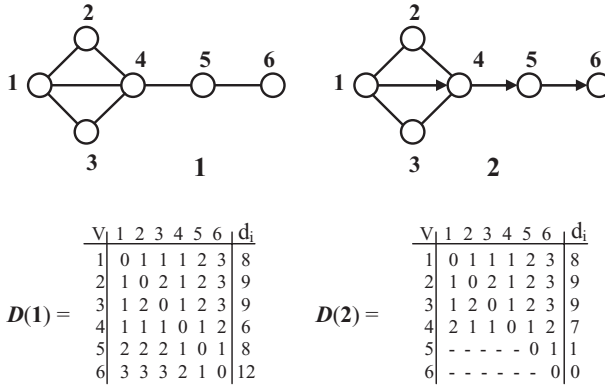
The average vertex distance (degree) $\langle d_i \rangle$ and average graph distance $\langle d \rangle$ (called also *graph radius* or *average path length* or *average degree of vertex-vertex separation*) are also defined:

$$\langle d_i \rangle = \frac{D}{V}; \quad \langle d \rangle = \frac{D}{V(V-1)} \quad (2.8a,b)$$

Examples illustrating distance matrix and derived descriptors are shown in Figure 5.4.

For graphs having loops, the denominator of Eq. (2.8b) changes to V^2 to include the diagonal elements of the distance matrix. *Distance degree distribution* $\{d_i\} \equiv \{d_1, d_2, \dots, d_V\}$, and *distance magnitude distribution* $\{d\} \equiv \{n_1, n_2, \dots, n_V\}$ are also defined from the distance matrix, where n_i is the frequency of occurrence of distance with magnitude i . *Vertex eccentricity*, e_i , is the maximum distance between vertex i and any of the remaining graph vertices. The largest vertex eccentricity is termed *graph diameter*. The vertex(es) with minimum eccentricity is defined as *graph center* [36]. An extended graph center definition [40,41] assumes the minimum eccentricity as a first criterion in a hierarchical series of criteria, which also includes the conditions for the minimum distance degree, and the minimum *distance degree sequence*, *DDS*. The latter is an ascending sequence of the distance magnitudes $1^{n_1} 2^{n_2} 3^{n_3} \dots (d_{max})^{n_{max}}$, with each distance frequency n_i as an exponent. An iterative vertex/edge centrality algorithm IVEC has been developed for the cases when the three hierarchical conditions do not suffice [42].

The distance degree distributions of graphs **1** and **2** are those given in the d_i columns of the matrices, whereas the distance magnitude distributions



$$D(G) = 52, \langle d_i \rangle = 8.67, \langle d \rangle = 1.73$$

$$D(DG) = 34, \langle d_i \rangle = 5.67, \langle d \rangle = 1.62$$

Figure 5.4. Distance matrices $D(1)$ and $D(2)$, total distances $D(1)$ and $D(2)$, average distance degrees $\langle d_i \rangle$, and average distances $\langle d \rangle$, of the undirected graph **1**, and the directed graph **2**, respectively.

of the two graphs are $\{d(G)\} \equiv \{14, 10, 6\}$ and $\{d(DG)\} \equiv \{11, 7, 3\}$, respectively. The vertex eccentricities in graph **1** are $e = 2$ for vertices 4 and 5, and $e = 3$ for the other four vertices. This specifies vertices 4 and 5 as graph centers according to the classical definition of Harary [36], and determines the graph diameter to be equal to 3. The extended graph center definition eliminates vertex 5, due to its larger distance degree (8 vs. 6), and leaves vertex 4 as a single graph center.

Several remarks should be made here related to the distances in directed graphs. Strictly speaking, directed graphs like graph **2** are disconnected, due to the lack of paths between some pairs of vertices, like the missing paths from vertex 6 to all other vertices. The distance between such pairs of vertices is equal to infinity, which makes the calculation of the total distance in directed graphs impossible. For practical purposes, one might discard such matrix entries as done in $D(2)$ above. However, as pointed out by Neuman *et al.* [43], the distance estimates produced in that way could be totally misleading. Indeed, in comparing the distance estimates for graphs **1** and **2**, e. g., $\langle d(2) \rangle = 1.62 < \langle d(1) \rangle = 1.73$, one may come to the wrong conclusion that the vertices in the directed graph **2** are closer to each other than those in the parent graph **1**. One way toward resolving these difficulties will be shown in Section 5. Another approach to the partial

disconnectedness of directed graphs was proposed by Newman *et al.* [43], who introduced the notion of strongly connected, as well as in- and out-component. A *strongly connected component* of a directed graph is a subgraph all vertices in which are connected by a finite path. The *out-component* contains vertices that can reach the strongly connected component but cannot be reached by any vertex of the strongly connected component. Conversely, the *in-component* contains all the vertices that cannot reach the vertices of the strongly connected component but can be reached from them. In the directed graph **2**, used in our examples, one can discern a strongly connected component formed by vertices 1-4, which can reach each other, as can be seen in the distance matrix $D(\mathbf{2})$ above. Vertices 5 and 6 form an in-component; they can be reached from the strongly connected component. The graph lacks an out-component.

Another feature of directed graphs is that the distance degrees d_i defined by Eq. (2.7a) as sums over the matrix row entries are in fact *distance out-degrees*, $d_i(out)$. The *distance in-degrees*, $d_i(in)$, which are obtained as sums over the distance matrix columns

$$d_i(in) = \sum_{j=1}^V d_{ij} \quad (2.7c)$$

are no more the same with their *out*-counterpart, because the directionality of graph arcs destroys the symmetry of the matrix. One may illustrate this point by comparing the two distributions for graph **2**: $\{d_i(\mathbf{2}, out)\} \equiv \{9, 9, 8, 7, 1, 0\}$ and $\{d_i(\mathbf{2}, in)\} \equiv \{12, 7, 4, 4, 4, 3\}$. Indeed, the total number of *in*- and *out*-distances in a directed graph must be equal. Vertices with large distance *out*-degrees may be of interest in the network analysis as important input nodes, whereas those with large distance *in*-degrees characterize essential output nodes.

Graph centers cannot be rigorously defined in directed graphs containing pairs of vertices with infinite distance between them. However, eliminating such vertices as potential graph centers, one may assess the remaining vertices with the same three criteria discussed above. In- and -out distances may define in principle different vertices as graph centers. In the example with the directed graph **2**, vertex 4 is classified as *out*-center by its minimum out-eccentricity value $e_4(out) = 2 = \min$ (vertices 5 and 6 are excluded from the competition of distance out-degrees). There is no competition for the *in*-center, which is in vertex 6, the only vertex that can be reached by all other vertices (all other vertices are excluded).

2.5. Weighted graphs

An essential generalization of the notion of graph, going beyond topology, enables the application of graph theory to every aspect of cellular networks. One may ascribe different vertex and edge weights, w_{ii} and w_{ij} , to match essential parameters of network species and their interactions. Vertex weights might characterize the level of expression of network species, as measured by mass-spectra, microarrays, HPLC, 2-D gel chromatography, and other methods. The edge weights in metabolic networks might characterize the enzymes expression. An edge weight in networks build of protein complexes denotes the number of proteins two complexes share. Other applications of weighted graphs exist or might be anticipated.

An edge or vertex *weight* could be any nonnegative natural number. (Weights having both positive and negative values has to be renormalized in order to enable using Eqs. (2.9-2.12). Weights can also be integers, as is the case with multigraphs, in which more than one edge connects some pairs of vertices. Another example is molecular networks, the different chemical nature of the atoms in which is sometimes labeled with vertex weights showing the number of their valence electrons. The *weighted adjacency matrix*, $\mathbf{WA}(G)$, has the edge weights w_{ij} as nondiagonal elements, and the vertex weights as diagonal elements, w_{ii} . All graph-invariants derived from the adjacency matrix of a directed or nondirected simple graph can be redefined for a weighted graph. Included here are the *weighted vertex degree*, w_i , and the corresponding *weighted vertex degree distribution*, $\{w_{max}, \dots, w_{min}\}$, *weighted adjacency*, $\mathbf{WA}(G)$, the *average weighted vertex degree*, $\langle w_i \rangle$, the *weighted connectedness*, $WConn$, the *weighted cluster coefficient*, wc_i , and the *weighted extended connectivity of order k*, ${}^k WEC$:

$$w_i = \sum_{j=1}^V w_{ij}; \quad \mathbf{WA}(G) = \sum_{i=1}^V \sum_{j=1}^V w_{ij} = \sum_{i=1}^V w_i \quad (2.9a,b)$$

$$\langle w_i \rangle = \frac{\mathbf{WA}}{V}; \quad WConn = \frac{\mathbf{WA}}{V^2} \quad (2.10a,b)$$

$$wc_i = \frac{\sum_{j \text{ adj } i} w_{ij}}{w_i(w_i - 1)} \quad (2.11)$$

$${}^n WEC = \sum_{i=1}^V {}^n w_i = \sum_{i=1}^V \sum_{j \text{ adj } i} {}^{n-1} w_j \quad (2.12)$$

3. How to Measure Network Complexity

3.1. Careful with symmetry!

There is a long-term controversy in the literature whether complexity of a structure increases with its connectivity or rather it passes through a maximum and goes down to zero for complete graphs. This is illustrated in Figure 5.5 with an example taken from Gell-Mann's book [44] “*The Quark and the Jaguar*”. The example includes two graphs with eight vertices; the first one is totally disconnected, whereas the second one is totally connected (complete) graph. It is argued that the two graphs are equally complex. The arguments in favor of this conclusion are based on the binomial distribution of vertex degrees in random graphs (Figure 5.6). Additional arguments in favor of such views come from Shannon's information theory [1]. According to it, the *entropy of information* $H(\alpha)$ in describing a message of N symbols, distributed according to some equivalence criterion α into k groups of $N_1, N_2, \dots, N_k$ symbols, is calculated according to the formula:

$$H(\alpha) = - \sum_{i=1}^k p_i \log_2 p_i = - \sum_{i=1}^k \frac{N_i}{N} \log_2 \frac{N_i}{N} \text{ bits/symbol} \quad (3.1)$$

where the ratio $N_i/N = p_i$ defines the probability of occurrence of the symbols of the i^{th} group.

In using Eq. (3.1) to characterize networks or graphs, it is the vertices that most frequently play the role of symbols or system elements. When the criterion of equivalence α is based on the orbits of the automorphism

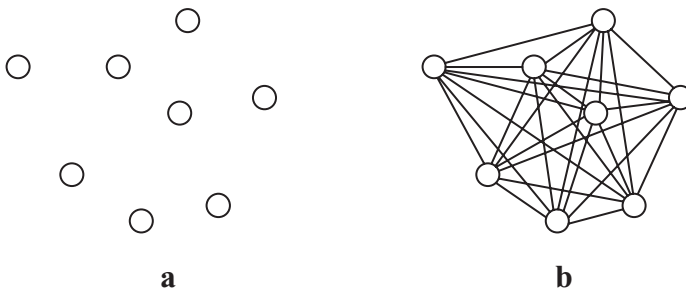


Figure 5.5. Which graph is more complex: the totally disconnected graph **a** or the complete graph **b**?

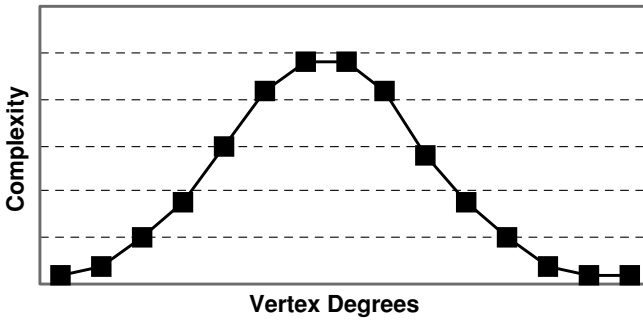


Figure 5.6. The binomial distribution of vertex degrees in random graphs is used as an argument that complexity of graphs passes through a maximum with the increase in connectivity.

group of the graph, all vertices of the totally disconnected graph belong to a single orbit, and the same is true for the vertices in the complete graph. Eq. (3.1) then shows that the information index $I(\alpha) = 0$ for both graphs. The same result is obtained when the partitioning of the graph vertices into groups is based on the equality of their vertex degrees, all of which are zeros in the totally disconnected graph, and all of which are of degree $N - 1$ in the complete graph.

The logic of the above arguments seems flawless. Yet, our intuition tells us that the complete graph is more complex than the totally disconnected graph. There is a hidden weak point in the manner the Shannon theory is applied, namely how symmetry is used to partition the vertices into groups. One should take into account that symmetry is a simplifying factor, but not a complexifying one. A measure of structural or topological complexity must not be based on symmetry. The use of symmetry is justified only in defining compositional complexity, which is based on equivalence and diversity of the elements of the system studied.

3.2. Can Shannon's information content measure topological complexity?

A different approach to characterizing structures by Shannon's theory was proposed in 1977 by Bonchev and Trinajstić in a study on molecular branching as a basic topological feature of molecules [15]. The approach was later generalized by constructing a finite probability scheme for a graph

[16]. Let the graph is represented by some kind of elements (vertices, edges, distances, cliques, etc.); let also assign a certain weight (value, magnitude) w_i to each of the N elements. Define the probability for a randomly chosen element i to have the weight w_i as $p_i = w_i / \sum w_i$, with $\sum w_i = w$, and $\sum p_i = 1$. The probability scheme thus constructed

Element	$1, 2, \dots, N$
Weight	$w_1, w_2, \dots, w_N$
Probability	$p_1, p_2, \dots, p_N$

enables defining a series of information indices, $I(w)$, with Shannon's Eq. (3.1).

Considering the simplest graph elements, the vertices, and assuming the weights assigned to each vertex to be the corresponding vertex degrees, one easily distinguishes the null complexity of the totally disconnected graph from the high complexity of the complete graph. The probability for a randomly chosen vertex i in the complete graph of V vertices to have a certain degree a_i is $p_i = a_i / A = 1 / V$, wherefrom Eq. (3.1) yields for the Shannon entropy of the vertex degree distribution the nonzero value of $\log_2 V$.

Our preceding studies [17, 45-47] have shown that a better complexity measure of graphs and networks is the vertex degree magnitude-based information content, I_{vd} . Shannon defines information as the reduced entropy of the system relative to the maximum entropy that can exist in a system with the same number of elements:

$$I = H_{\max} - H \quad (3.2)$$

The Shannon entropy of a graph with a total weight W and vertex weights w_i is given by a formula derived from Eq. (3.1):

$$H(W) = W \log_2 W - \sum_{i=1}^V w_i \log_2 w_i \quad (3.3)$$

The maximum entropy is obtained when all $w_i = 1$:

$$H_{\max} = W \log_2 W \quad (3.4)$$

From Eqs.. (3.2-3.4), substituting also $W = A$ and $w_i = a_i$, one obtains the equation for the information content of the vertex degree distribution of a graph, I_{vd} :

$$I_{vd} = \sum_{i=1}^V a_i \log_2 a_i \quad (3.5)$$

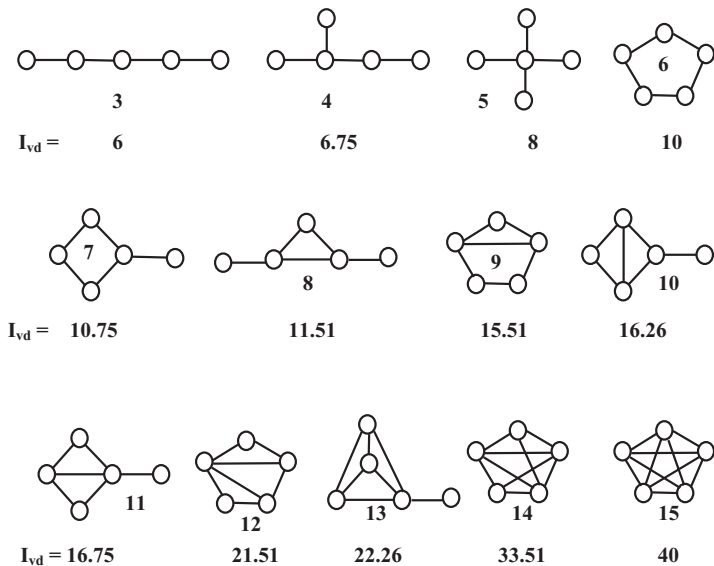


Figure 5.7. Thirteen graphs with five vertices ordered according to their increasing complexity, adequately matched by the values of the information index for the vertex degree distribution.

The analysis has shown that the I_{vd} index satisfies the criteria for a complexity measure and can be recommended for assessments of network complexity [17, 45-47]. It increases with the connectivity and other complexity factors, such as the number of branches, cycles, cliques, etc., as shown in the series of graphs in Figure 5.7. The increase in the number of branches increases the complexity index, as seen in the sequences of graphs $3 \rightarrow 4 \rightarrow 5$, $6 \rightarrow 7 \rightarrow 8$, $9 \rightarrow 10$, and $12 \rightarrow 13$. The number of cycles is a considerably stronger complexity factor, as demonstrated in the sequence of graphs with one to five cycles: $6 \rightarrow 9 \rightarrow 12 \rightarrow 14 \rightarrow 15$.

3.3. Global, average, and normalized complexity

A variety of graph-invariants have been examined as measures of topological complexity [48-50]. Since they are directly applicable to networks, we shall review some of the most promising ones, systematizing them in a scheme discussed below.

A series of connectivity descriptors was introduced in Section 2.2. Total adjacency A is the count of all pairwise neighborhood relationships, $a_{ij} = 1$, each of which denotes a link directed from vertex i to vertex j . Total

adjacency is thus equal to the total number of directed edges in the graph. In nondirected graphs, one usually equalizes total adjacency to the doubled number of edges, $A = 2E$. Each nondirected edge $\{ij\}$ in these graphs is in fact an abbreviated notation for two directed edges, one from i to j , and the second one from j to i , respectively. One might then abandon the tradition, and use the symbol E for the total number of (directed, in- and out-) edges in both directed and nondirected graphs, i. e., to use E for the total number of nonzero adjacency matrix entries a_{ij} . We may summarize this analysis by interpreting the redefined total adjacency A as a first level topological complexity measure, and term it graph (or network) *global edge complexity*, E_g

$$A = \sum_{i=1}^V \sum_{j=1}^V a_{ij} = \sum_{i=1}^V a_i = E_g \quad (3.6)$$

A similar reinterpretation may be made to the average vertex degree $\langle a_i \rangle$, and connectedness, $Conn$, introduced by Eq. (2.4b). One may call the average vertex degree thus defined *average edge complexity*, E_a , the averaging being defined per vertex. On its turn, connectedness can be regarded as *normalized edge complexity*, E_n , because it is redefined as the ratio of the global edge complexity $E_g = A = E$ and the number of edges in the complete graph with loops at each of its vertices, $E(K_V)$:

$$\langle a_i \rangle = \frac{A}{V} = \frac{E_g}{V} = E_a ; \quad Conn = \frac{A}{V^2} = \frac{E_g}{V^2} = E_n \quad (3.7a,b)(19a,b)$$

When the graph contains no loops, the denominator of Eq. (3.7b) may be replaced by the $V(V-1)$, eliminating thus the potential contributions from the adjacency matrix diagonal elements of the complete graph.

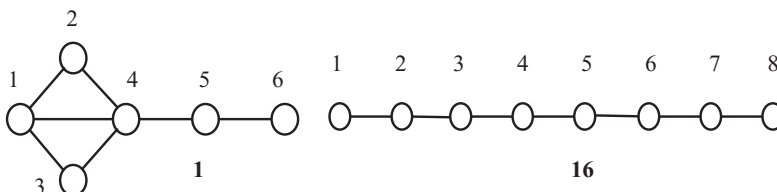
We have thus presented three individually introduced connectivity descriptors, as three versions of the simplest topological complexity measure: the global, average, and normalized edge complexity. We shall use this triple scheme in presenting other, more sophisticated measures of network complexity. Such more advanced complexity indices are needed because connectedness (the relative edge complexity) is a descriptor that counts only the total number of vertex interconnections, but does not account for the specific way these connections occur. At the same connectedness two networks could differ in their complexity by orders of magnitude. It may be anticipated that the global measures will be of major use in characterizing

pathways and small networks, whereas the large networks will be better assessed by the average and relative complexity measures.

3.4. The subgraph count, SC , and its components

What would be the next step in the search for more adequate network complexity measures? We started in the preceding subsection with counting the simple subgraphs, the edges, and called this descriptor edge complexity. It seems logical to continue with counting the subgraphs containing two edges. The importance of the two-bonds molecular fragments for the properties of chemical compounds has been early understood, and the total number of these fragments is known in chemical theory as Platt's index [51]. Bertz used this index as a measure of molecular complexity [17], calling the two-edge fragments "connections". He also constructed an information complexity measure proceeding from the distribution of the two-edge subgraphs into equivalence groupsb [18]. The Platt index is considerably better complexity measure than the number of edges. At the same number of edges the Platt index increases rapidly with the presence of complexifying factors like branches and cycles.

Such an example is shown in Figure 5.8, in which graph **1** having two cycles is compared to the path graph **16** having the same number of seven edges. The number of two-edge subgraphs is denoted as 2SC , meaning



Graph **1**: 124, 134, 142, 143, 145, 213, 214, 243, 245, 314, 345, 456
 $E = 7$, ${}^2SC = 12$, ${}^2SC_a = 2$, ${}^2SC_n = 0.5$, $Conn = {}^1SC_n = 0.233$

Graph **16**: 123, 234, 345, 456, 567, 678
 $E = 7$, ${}^2SC = 6$, ${}^2SC_a = 0.75$, ${}^2SC_n = 0.036$, $Conn = {}^1SC_n = 0.125$

Figure 5.8. The larger complexity of graph **1** as compared to graph **16** is demonstrated by the total, average and normalized number of two-edge subgraphs 2SC , 2SC_a , and 2SC_n , respectively, as well as by the graph connectedness $Conn$, which is identical to the normalized number of edges, 1SC_n .

2^{nd} -order subgraph count (*vide infra*). The corresponding average and relative substructure counts of 2^{nd} -order are also shown.

The two graphs differ considerably by their complexity, because the path graph **16** lacks any complexifying structural features, whereas graph **1** incorporates two cycles. Connectedness, *Conn*, does not reflect to a sufficient degree this difference in complexity of the two graphs ($\text{Conn}(\mathbf{1}) : \text{Conn}(\mathbf{16}) = 1.9$), whereas the normalized two-edge complexity 2SC_n of graph **1** is shown to be much higher than that of **16** ($0.5 : 0.036 = 13.9$).

In calculating the 2SC_n values:

$${}^2SC_n = \frac{{}^2SC}{{}^2SC(K_V)} \quad (3.8)$$

we made use of the formula derived [52] for the 2^{nd} -order subgraph count of the complete graph K_V :

$${}^2SC(K_V) = E \times (a_i - 1) = \frac{1}{2} V(V-1)(V-2) \quad (3.9)$$

The analysis performed in chemical graph theory has shown that the Platt index still fails to mirror some complexity structural patterns, and the search for better measures has continued. A next logical step would be to use the number of three-edge subgraphs, 3SC . Such an index has been used in chemical graph theory as Gordon-Scantlebury index [53], however, it has not been tested as a complexity measure. Instead, Bertz and Herndon proposed in 1986 the idea to use the total subgraph count, *SC*, which includes subgraphs of all sizes, including the graph itself, regarded as a proper subgraph [54]. The idea remained unused until the late 1990s, when Bertz [26,27] and Bonchev [9, 24, 25, 28, 29] independently and simultaneously developed the approach in detail. Bertz applied the *SC* global index to the synthesis planning in organic chemistry, while the present author derived explicit *SC* formulae for some basic classes of graphs, and the represented the total subgraph count as an ordered set of counts of subgraphs having a given number of edges. The set $\{SC\}$ begins with the number of vertices V , regarded as null-order index, 0SC , followed by the number of edges E , as first-order index, 1SC , the two-edge subgraphs, as the second-order index, 2SC , etc.:

$$SC = {}^0SC + {}^1SC + {}^2SC + \dots + {}^E SC \quad (3.10a)$$

$$\{SC\} = \{{}^0SC, {}^1SC, {}^2SC, \dots, {}^E SC\} \quad (3.10b)$$

Illustrating the formulas, one obtains for graph **1** the total subgraph count $SC = 90$, and the set of its null- through seventh-order terms $\{SC\} = \{6, 7, 12, 20, 22, 16, 6, 1\}$. The calculations were performed with the program SUBGRAU developed by Rücker and Rückerm [55].

In assessing the complexity of large networks, formulas (22a,b) lead to combinatorial explosion. By this reason, one might recommend using for such purposes only the first-, second-, and third-order subgraph count, whereas the higher orders and the total count could be calculated for pathways and small subnetworks. It is worth mentioning that connectedness (or connectance), which is used almost exclusively in characterizing dynamic networks, appears naturally as the normalized first-order term in the series (22a,b). One might anticipate a broader application of the higher terms, particularly 2SC_n and 3SC_n , due to their much higher sensitivity to the complexifying details of the networks. For the normalizing of these terms one may use the formulas we derived for the three-edge subgraph count 3SC of the complete graph K_V , as well as for its components, the counts of triangular, linear, and star type three-edge subgraphs:

$${}^3SC(K_V) = \frac{1}{6}V(V-1)(V-2)(4V-11) \quad (3.11)$$

$${}^3SC(K_V, \text{triangle}) = \frac{1}{6}V(V-1)(V-2) \quad (3.12)$$

$${}^3SC(K_V, \text{linear}) = \frac{1}{2}V(V-1)(V-2)(V-3) \quad (3.13)$$

$${}^3SC(K_V, \text{star}) = \frac{1}{6}V(V-1)(V-2)(V-3) \quad (3.14)$$

The comparison of the third-order subgraph counts of graphs **1** and **3**, 20 vs. 5, shows again a considerably higher complexity of graph **1** as compared to the assessment based on the graph connectedness (connectance). One may also recommend to use for more detailed characterization of complex networks, the separate counts of the three kinds of three-edge subgraphs – triangles, stars, and linear ones, 3SC_t , 3SC_s , and 3SC_l , which were previously shown to produce high correlations with physicochemical properties [56].

3.5. Overall connectivity, OC

The subgraph count presentation as an ordered set of components with increasing size may be regarded as a part of a more general scheme [57]. The latter defines a certain *overall* graph-invariant X , by the sum over the values this invariant has for each of the subgraphs. Also, the contributions of all subgraphs having k edges are combined in single term, kX . An ordered set $\{X\}$ on all k -terms is also constructed, and the initial terms $k = 0, 1, 2, 3, \dots$, called null-, first-, second-, etc. order terms, can be independently used to characterize the graph properties.

$$X = \sum_{k=1}^E {}^kX; \quad \{X\} = \{{}^0X, {}^1X, {}^2X, \dots, {}^EX\} \quad (3.15)$$

In addition, one can also define the average value of X per vertex, X_a , as well as its normalized value, $0 \leq X_n \leq 1$:

$$X_a = \frac{X}{V}; \quad {}^kX_a = \frac{{}^kX}{V} \quad (3.16a,b)$$

$$X_n = \frac{X}{X(K_V)}; \quad {}^kX_n = \frac{{}^kX}{{}^kX(K_V)} \quad (3.17a,b)$$

The scheme can be further detailed by using within each kX term the counts of subgraphs of different topology, e.g., for three edge subgraphs the counts of triangles, stars, and linear (or path) graphs [56].

The simplest graph-invariant that can be incorporated into this scheme is the subgraph count, SC , as shown in the foregoing. The next basic candidate is the graph adjacency A , defined by Eq. (2.2b). By summing up the adjacencies of all k^{th} -order subgraphs kG_i , with $k = 0, 1, 2, 3, \dots, E$, one defines [28,29] the *overall connectivity* $OC(G)$ of the graph G :

$$OC(G) = \sum_{k=1}^E {}^kOC = \sum_{k=1}^E \sum_i {}^kA_i ({}^kG_i \subset G) \quad (3.18a)$$

$$\{OC\} = \{{}^0OC, {}^1OC, {}^2OC, \dots, {}^EOC\} \quad (3.18b)$$

Equations. (3.18a,b) yield for graph **1** the overall connectivity value $OC = 936$, and the set of its 0- to 7-th order terms: $\{OC\} = \{14, 38, 101,$

210, 264, 212, 83, 14}. It should be mentioned that in the first publications defining overall connectivity [24,25], the latter was termed *topological complexity* and denoted by TC . This name was later changed [28,29] to overall connectivity to account for the fact that this is not the only measure of topological complexity.

According to the general scheme, the overall connectivity index can also be presented as averaged per vertex, and in a normalized form. To facilitate the calculation of the first-, second-, and third-order normalized index, Eqs. (3.19)-(3.21) were derived, along with Eqs. (3.22)-(3.24) for the three different topological shapes of the three-edge subgraphs:

$${}^1OC(K_V) = V(V-1)^2 \quad (3.19)$$

$${}^2OC(K_V) = \frac{3}{2}V(V-1)^2(V-2) \quad (3.20)$$

$${}^3OC(K_V) = \frac{1}{6}V(V-1)^2(V-2)(16V-45) \quad (3.21)$$

$${}^3OC(K_V, \text{triangle}) = \frac{1}{2}V(V-1)^2(V-2) \quad (3.22)$$

$${}^3OC(K_V, \text{linear}) = 2V(V-1)^2(V-2)(V-3) \quad (3.23)$$

$${}^3OC(K_V, \text{star}) = \frac{2}{3}V(V-1)^2(V-2)(V-3) \quad (3.24)$$

The overall topological indices scheme, defined by Eqs. (3.15-3.17), has also been applied to other graph invariants, such as the Wiener number [58-60] and the Zagreb indices [56,61,62]. These overall indices have also shown properties of complexity measures.

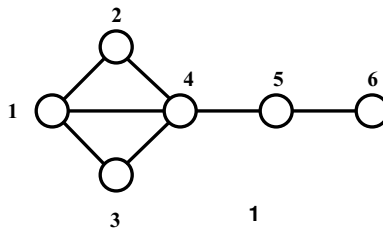
3.6. The total walk count, TWC

Rücker and Rücker have proposed [30,31] a similar scheme for assessing the graph complexity by the *total walk count*, TWC . This complexity measure is obtained by counting all walks ${}^l w_i$ of all lengths l , the maximum

walk length being limited by the graph size:

$$TWC = \sum_{l=1}^{V-1} {}^lWC = \sum_{l=1}^{V-1} \sum_i {}^l w_i \quad (3.25)$$

For graph **1**, one finds $TWC = 1154 \{14, 38, 100, 272, 730\}$. The length-one walks are just the doubled number of edges, since each of the two ends of an edge is used as a walk starting point. There are two types of walks of length two: forward and back along the same edge ($1 \rightarrow 2 \rightarrow 1$) and forward along two adjacent edges ($1 \rightarrow 2 \rightarrow 4$). Each of these two types then generates two different types of walks of length three, with the third step backside ($1 \rightarrow 2 \rightarrow 1 \rightarrow 2$; $1 \rightarrow 2 \rightarrow 4 \rightarrow 2$) or along a different edge ($1 \rightarrow 2 \rightarrow 1 \rightarrow 4$; $1 \rightarrow 2 \rightarrow 4 \rightarrow 3$), etc.



Scheme 5.1.

The number of walks of length l , is obtained from the l^{th} power of the adjacency matrix. For calculating the normalized lWC_n indices, one has to use Eq. (3.26) derived for the respective value in the complete graph with the same number of vertices. One would then find for graph **1**, ${}^2WC_n = 0.253$ and ${}^3WC_n = 0.133$.

$${}^lWC(K_V) = V(V-1)^l \quad (3.26)$$

Like the subgraph count and the overall connectivity, the total walk count is an adequate measure of graph complexity, showing patterns of regular increase with the graph size, connectedness, and the basic structure complexifying factors such as the number, size and the kind of interconnectedness of the graph cycles and branches [31]. Figure 5.9 illustrates these conclusions, providing the same ordering of increasing complexity of graphs **3** to **15** like the one produced in the foregoing by the I_{vd} index. The complexity measures discussed in Section 3 have all been previously published. In the next Section 4, we report some new developments.

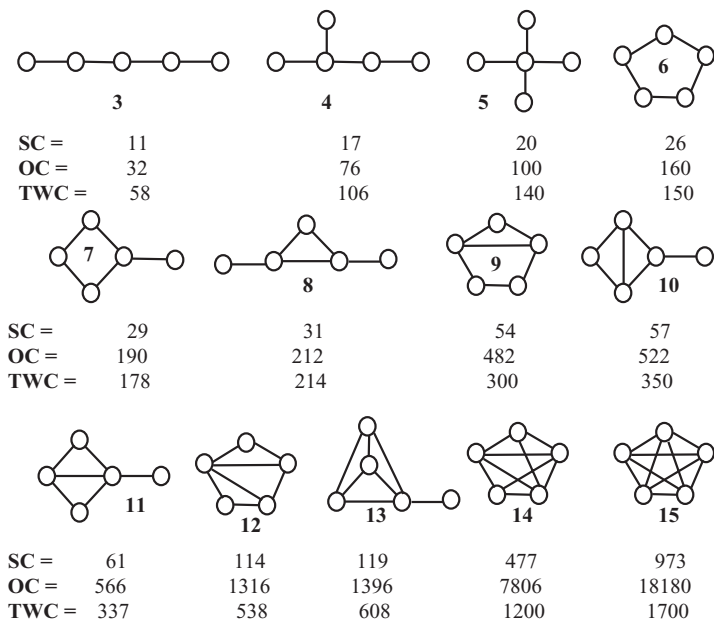


Figure 5.9. Thirteen graphs with five vertices ordered according to their increasing complexity, adequately matched by the values of the subgraph count SC , overall connectivity OC , and the total walk count TWC .

4. Combined Complexity Measures Based on the Graph Adjacency and Distance

4.1. The A/D index

Networks with high complexity are characterized by both high vertex-vertex connectedness and small vertex-vertex separation (the small-world concept of Watts and Strogatz [63]). Therefore, it seems logical to use both quantities in characterizing network complexity. The ratio $A/D = \langle a_i \rangle / \langle d_i \rangle$ of the total adjacency and the total distance of the graph or, equivalently, the ratio of the average vertex degree $\langle a_i \rangle$ and the average distance degree $\langle d_i \rangle$, may be regarded as a logical approach to such a complexity measure. At a constant number of vertices, the A/D index has a minimum value in path graphs, P_V , which are characterized by low connectivity and long distances. In contrast, the A/D ratio has a maximum value in the complete graphs, K_V , which are maximally connected and all

of their vertices have only a unit distance separation. The classes of star graphs, S_V , and monocyclic graphs, C_V , are of intermediate complexity and their A/D indices are between these two extremes.

$$A/D(P_V) = \frac{2(V-1)}{V(V-1)(V+1)/3} = \frac{6}{V(V+1)} \quad (4.1)$$

$$A/D(K_V) = \frac{V(V-1)}{V(V-1)} = 1 \quad (4.2)$$

$$A/D(S_V) = \frac{2(V-1)}{2(V-1)^2} = \frac{1}{V-1} \quad (4.3)$$

$$A/D(C_V, \text{odd}) = \frac{2V}{2V(V^2-1)/8} = \frac{8}{(V^2-1)} \quad (4.4a)$$

$$A/D(C_V, \text{even}) = \frac{2V}{V^3/4} = \frac{8}{V^2} \quad (4.4b)$$

As shown in Eq. (4.2), the A/D index of the complete graph is equal to a unity; therefore, all graphs have their A/D values within the 0 to 1 range. Like all normalized complexity indices this index decreases rapidly with the graph size for path graphs, monocyclic graphs, and other weakly connected graphs, the distance in which dominates strongly over adjacency. Some degeneracy of the index (having two or more nonisomorphic graphs with the same A/D ratio) should be expected, because both the total adjacency A and the total distance D are degenerate. What might be a more serious problem is the insensitivity to some more subtle topological features of branching and cyclicity, which sometimes produces incorrect assessments of graph complexity (See Table 5.1, and the examples in the next subsection). Yet, the fine details of topological structure might be inessential when dealing with large networks, for which the A/D index could prove to be a sufficiently accurate measure of structural complexity. For smaller subnetworks and particularly pathways, perhaps a better recommendation would be to make use of the new structural index presented in Section 4.2.

Table 5.1. The newly defined complexity index B matches well the complexity ordering of six-vertex graphs with the same connectedness (Fig. 5.4) as produced by four other complexity measures

Graph	A/D	$B = \Sigma a_i/d_i$	SC	OC	TWC	I_{vd}
17	0.250	1.833	62	535	852	15.06
18	0.231	1.636	56	475	754	14.75
19	0.231	1.567	52	426	598	14.00
20	0.231	1.464	43	329	450	12.75
21	0.222	1.558	53	444	708	13.75*
22	0.222	1.544	49	394*	662	14.00
23	0.222	1.483	49	396*	556	13.51
24	0.222	1.464	37	264	372	12.00
25	0.214	1.439	44*	343*	564	13.51
26	0.214	1.417	48*	386*	540	13.51
27	0.207	1.408	45	354	602	13.51
28	0.207	1.354	42	318	480	12.75
29	0.194	1.260	37	266	490	12.75

*The three pairs of values denoted by asterisks indicate the few cases in which the new complexity index B produced inversed graph ordering as compared with the ordering resulting from the four known complexity measures.

4.2. The complexity index B

The ratio $b_i = a_i / d_i$ of the vertex degree a_i and its distance degree d_i is a local invariant with interesting centric properties. It is ≤ 1 , the equality occurring for the central vertex in the star graphs, as well as for every vertex in the complete graph. The sum over the b_i values of all graph vertices may be expected to behave similarly to the A/D ratio, with less degeneracy, and more sensitivity to local topology. We define this sum as a new complexity index B :

$$B = \sum_{i=1}^V \frac{a_i}{d_i} \tag{4.5}$$

Several equations derived for the b_i and B indices shed some light on the properties of these complexity descriptors. In complete graphs, K_V , in which $a_i = d_i = V-1$, and $b_i = 1$ for every vertex, the B index is simply equal to the number of vertices V :

$$B(K_V) = V \tag{4.6}$$

In star graphs, S_V , in which the central vertex c is of degree $V-1$, and all other vertices are terminal (t) with degree 1, one obtains

$$b_t = \frac{1}{2V-3}; \quad b_c = 1; \quad B(S_V) = \frac{3V-4}{2V-3} \quad (4.7a,b,c)$$

In (mono)cyclic graphs, C_V , all vertices have degree two, and have the same distance degree. The expression for the latter differs slightly for the odd- and even-membered cycles:

$$C_V(odd): \quad b = \frac{8}{V^2-1}; \quad B = \frac{8V}{V^2-1} \quad (4.8a)$$

$$C_V(even): \quad b = \frac{8}{V^2}; \quad B = \frac{8V}{V^2} = \frac{8}{V} \quad (4.8b)$$

The B index values begin at $B = 3$ for the odd-membered cycles and at $B = 2$ for the even-membered cycles, and gradually decrease with the cycle size to the zero limit at $V \rightarrow \infty$.

In the path graphs, P_V , the two terminal vertices are of degree 1 and all others are of degree two. The formulas for the local b_i indices depend on the position $i = 1, 2, 3, \dots, V$ of the vertex, counting from the end of the chain. Different equation is obtained only for the central one or two vertices c :

$$b_i = \frac{2a_i}{V^2 - (2i-1)V + 2i(i-1)} \quad (4.9a)$$

$$b_c(odd) = \frac{8}{V^2-1}; \quad b_c(even) = \frac{8}{V^2} \quad (4.9b)$$

No closed form equation can be obtained for the B index of path graphs. However, the presence of the V^2 term in the denominator of the local b_i and b_c indices shows that at large path length they, as well as the B index, will tend to zero considerably faster than the respective indices for the monocyclic graphs, which decrease with V only linearly.

The testing of the new complexity measure with graphs **3 – 15**, used in Section 3 to demonstrate the behavior of other complexity measures, has shown a perfect match with the ordering produced by the subgraph count, overall connectivity, total walk count and the information on the vertex degree distribution (Figure 5.10).

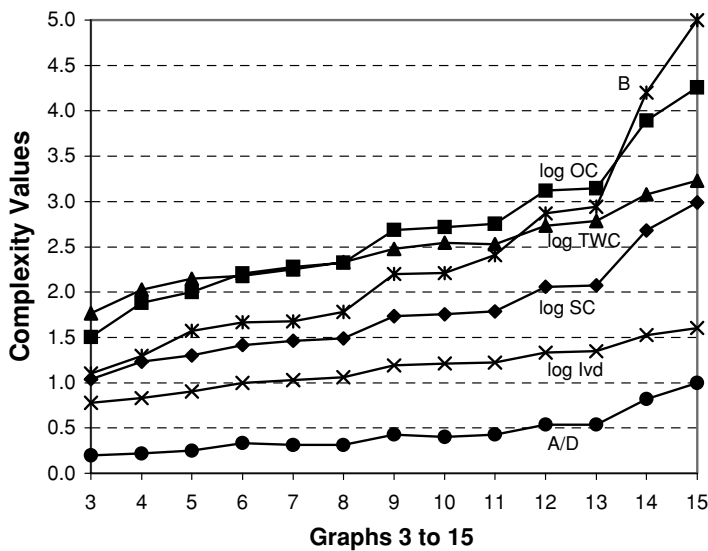


Figure 5.10. With few exceptions for the A/D and I_{vd} indices all the six complexity measures match the increase in complexity of graphs 3 through 15.

The A/D index also captured the basic complexity features in this series of graphs to increase with the number of branches and cycles. However, it is less sensitive to subtle details of graph topology, which resulted in three inverse orderings and three degeneracies.

B ordering: **3**(1.105) → **4**(1.294) → **5**(1.571) → **6**(1.667) → **7**(1.677) → **8**(1.783) → **9**(2.200) → **10**(2.211) → **11**(2.410) → **12**(2.867) → **13**(2.943) → **14**(4.200) → **15**(5.000)

A/D ordering: **3**(0.200) → **4**(0.222) → **5**(0.250) → **7**(0.313) = **8**(0.313) → **6**(0.333) → **10**(0.400) → **9**(0.429) = **11**(0.429) → **12**(0.538) = **13**(0.538) → **14**(0.818) → **15**(1.000)

Additional comparisons between the new A/D and B indices and the four selected known complexity measures are shown in Table 5.1 for the 13 six-vertex graphs from Figure 5.11. Once again, the B index captures the complexity features of the graphs examined much better than the A/D

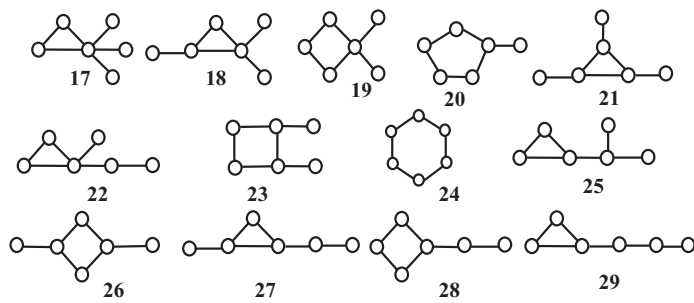


Figure 5.11. Thirteen graphs with six vertices and six edges used as a test for the sensitivity of the complexity measures.

ratio. The A/D index not only shows high degeneracy but in the degenerate quartet and triplet of graphs it produces the same complexity estimate for graphs that all other five indices distinguish drastically, e.g., **18** and **20**, **21** and **24**, and others. The *B* index generates the same ordering as the total walk count *TWC*, and has minimal number of reorderings (denoted by asterisks in Table 5.1) with the subgraph count *SC*, the information index for the vertex degree distribution I_{vd} , and the overall connectivity index *OC*, the latter four indices not producing identical orderings as well. The *B* index has also a single degeneracy, slightly worse than *OC* and *TWC* with no degeneracy, and better than *SC* with two, and I_{vd} with even six degenerate values. All this characterizes the index *B* introduced here as a convenient measure of graph complexity, a measure that shows similar behavior to other well established and sensitive complexity measures, and does not require substantial computational time.

5. Vertex Accessibility and Complexity of Directed Graphs

In Section 2.4 we have discussed the misleading results that are obtained for the graph radius (the average path length or the average graph distance) in directed graphs when one simply neglects the infinite distances between the pairs of vertices for which no path exists, and averages the remaining distances. Such calculations would produce the false impression that the radius of directed graphs is smaller than that of the parent undirected graph. A correcting procedure that restores the normal distance ratios between

the parent undirected graph and the directed graphs generated from it was recently described [45]. It introduces a parameter called *vertex accessibility*, $Acc(DG)$, which accounts for the degree to which the vertices in directed graphs are mutually accessible via finite paths. The vertex accessibility of a directed graph DG is defined as the ratio of the number of finite distances in the directed graph, $N_d(DG)$, and the total number of distances in the parent undirected graph $N_d(G)$:

$$Acc(DG) = \frac{N_d(DG)}{N_d(G)} \quad (5.1)$$

In Eq. (5.1), $N_d(G) = V^2$ (the squared total number of vertices V) in the general case of connected undirected graphs with loops. In that case, $N_d(DG)$ includes also the number of loops, as given with all $d_{ii} = 1$ appearing in the main diagonal of the distance matrix. If no loops can in principle exist in a certain type of networks, then $N_d(G) = V(V-1)$ should be used.

Equation (5.1) enables obtaining a more realistic estimate of the average path length $\langle d \rangle$ in a directed graph. Dividing $\langle d \rangle = D/N_d$, by the vertex accessibility, one normalizes this quantity to the case of complete vertex accessibility. The *adjusted average distance (adjusted average path length)*, $AD(DG)$:

$$AD(DG) = \frac{\langle d \rangle}{Acc} = \frac{D \times V^2}{N_d^2} \quad (5.2)$$

thus defined, is larger than the average distance in the parent undirected graph, and can be used for comparisons of the average degree of separation in directed graphs. As in Eq. (5.1), for DGs without loops V^2 can be replaced by $V(V-1)$. The calculation made for the vertex accessibility of directed graph **2** (see the distance matrix of this graph in Section 2.4) produces $ACC(2) = 21/(6 \times 5) = 0.7$. From here, with Eq. (5.1) one obtains for the adjusted average distance of this graph, $AD(2) = 1.62/0.7 = 2.31$. Thus, the unrealistic value of 1.62, after the adjustment turned from smaller to considerably larger than the corresponding value of 1.73 for the parent undirected graph 1.

Vertex accessibility can also be used to define a more realistic measure of the connectedness of directed graphs. The new measure might be termed *accessible connectedness*, $AConn(DG)$:

$$AConn(DG) = Conn(G) \times Acc(DG) = Conn(G) \times \frac{N_d(DG)}{N_d(G)} \quad (5.3)$$

Illustrating Eq. (5.3), the calculation for the directed graph **2** results in $AConn(2) = 0.214$, down from the unadjusted value of $Conn(2) = 0.306$ calculated in Section 2.2, a value that was unrealistically close to that of the parent undirected graph $Conn(1) = 0.389$.

Similar adjustment may be made to the A/D index of directed graph. Substituting the misleading distance D by its adjusted counterpart AD , one defines the A/AD complexity measure of directed graphs.

Some classes of directed graphs are of interest, because of the special relations existing for their vertex accessibility and the adjusted indices derived from it. Such is the special class in which all edges are directed and their direction is the same (all linear or clockwise, etc.). It can be easily shown that for monocyclic and complete graphs of this class, there is a complete accessibility of all vertices, at the cost of considerably larger average path length than that of the parent undirected graph. Thus, the directed graph DC_6 has a total distance of 90, a vertex distance of 15, and an average distance of 3, whereas its parent undirected graph C_6 has a total distance of 54, a vertex distance of 9, and an average distance of only 1.8. The directed graph DK_5 has a total distance of 30, a vertex distance of 6, and an average distance of 1.5, as compared to the parent complete graph K_5 having a total distance of 20, a vertex distance of 4, and an average distance of 1.

The directed path graph and star graph shown in Figure 5.12 do not have complete vertex accessibility. The actual accessibility, the adjusted average distance, and the adjusted connectedness can be assessed by the following

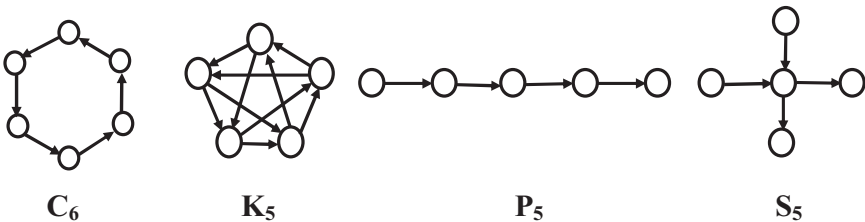


Figure 5.12. Special subclasses of directed graphs belonging to the classes of monocyclic, complete, path, and star graphs, respectively. The DC_V and DK_V subclasses shown have a complete vertex accessibility. Directed star graphs DS_V have the highest accessibility when a half of the arcs are incoming to and the other half of the arcs are outgoing from the central vertex.

formulae:

$$\begin{aligned} Acc(DP_V) &= \frac{1}{2}; & AD(DP_V) &= \frac{2(V+1)}{3}; \\ AConn(DP_V) &= \frac{1}{2V} \end{aligned} \quad (5.4a,b,c)$$

$$Acc(DS_V, odd) = \frac{V+3}{4V}; \quad Acc(DS_V, even) = \frac{V_2 + 2V - 4}{4V(V-1)} \quad (5.5a,b,)$$

$$\begin{aligned} AD(DS_V, odd) &= \frac{8V(V+1)}{(V+3)^2}; & AD(DS_V, even) &= \\ &= \frac{8V(V-1)(V^2-2)}{(V^2+2V-4)^2} \end{aligned} \quad (5.6a,b,)$$

$$AConn(DS_V, odd) = \frac{V+3}{4V^2}; \quad AConn(DS_V, even) = \frac{V^2 + 2V - 4}{4V^2(V-1)} \quad (5.7a,b,)$$

6. Complexity Estimates of Biological and Ecological Networks

Networks are universal means for analyzing systems in their entirety, and for capturing the systems complexity patterns [64]. Not surprisingly, after the revolution in network theory started [65] in 1999, and the focus has shifted from random networks to dynamic evolutionary ones [66] up to a half of all working papers of the Santa Fe Institute of Complexity have been devoted to networks [67]. The physical nature of the network nodes and their interactions is inessential in this analysis. In biological networks nodes can represent proteins [68-71] or protein complexes [72], genes [73-75], metabolites [76-78], neurons [79], etc. The type of “interaction” that connects two nodes in the network in an edge or arc could also vary from chemical binding to regulatory effects to signal transduction to nerve impulse. There are also networks in which there is no real interaction but the

edge may stand, for example, for the presence of the same species (proteins or genes) in different complexes. In food webs, the nodes represent different kind of biological species, while the type of interaction is “who is eating whom”. However different systems the networks may represent, they all have common features and share common structural patterns based on the connectivity of their constituents. Complexity measures make possible the characterization of these common network features in a general quantitative scale, providing thus the means for comparisons and quantitative evolutionary models.

6.1. Networks of Protein Complexes

Proteins tend to associate with each other forming complexes. The size of these complexes may vary within a rather broad range. Figure 5.13 presents the network of protein complexes taking part in the RNA metabolism of *Saccharomyces Cerevisiae* (data taken from Gavin *et al.* [72]). The 28 complexes contain 692 proteins, which amounts in average to almost 25 proteins in a complex, the actual sizes ranging from 2 to 133 complexes. The complexes are denoted by sequential numbers as given in the Supplementary Table 5.3 of the data source [72]. Each edge in Figure 5.13 stands for sharing proteins between the corresponding two complexes. The exact number of shared proteins is not shown as edge weights, due to the graph complexity. In the majority of cases the pairs of complexes share only one protein. In four cases, the number of shared proteins is between ten and fifteen. The calculations of the complexity measures of this weighted undirected graph are also performed for the basic topology of the parent non-weighted graph.

The graph actually shows the giant component (a term used to denote the graph component that incorporates the majority of vertices) of the network, the latter also containing three complexes that not share proteins with other complexes. The giant component is highly connected with a 106 non-weighted edges or basic adjacency of 212. This leads to average basic vertex degree of 8.48, and connectedness of 0.353. The corresponding values based on the edge weights are: weighted vertex adjacency of 1124, average weighted vertex degree of 44.96, and weighted connectedness of 1.87. This high connectedness evidences for the high stability against attacks or mutations, and indicates the importance of the RNA metabolism for the cell survival. High adjacency/connectedness values are obtained also

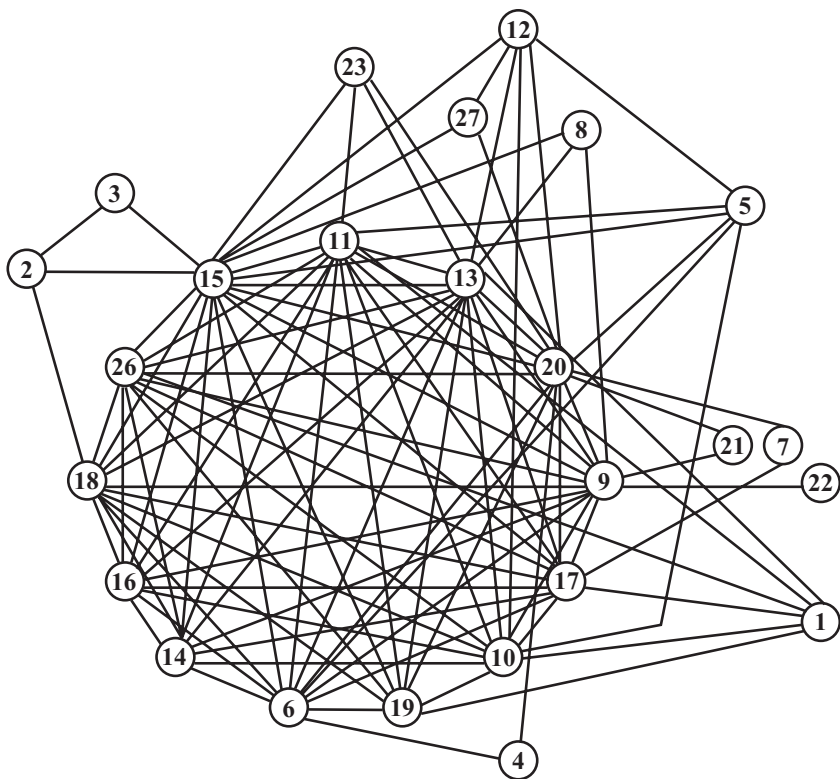


Figure 5.13. The network of the protein complexes functional group of RNA metabolism in *Saccharomyces Cerevisiae*. The complexes sequential numbers and connectivity table are those from Gavin *et al.* [72]. A pair of vertices are connected by an edge when they share at least one protein. (Not shown are three complexes that do not share any proteins). The high complexity of the network indicates the high stability of the RNA metabolism against random attacks and mutations.

for the networks of protein complexes responsible for transcription/DNA maintenance/chromatin structure, and for protein synthesis and turnover (Table 5.2). The comparison of the connectivity descriptors in Table 5.2 also allows concluding that the biological functions of signaling, cell cycle, and cell polarity and structure are more vulnerable against such attacks. Similar conclusions can be drawn from Table 5.3, which presents the values of the more recent complexity measures calculated for the weighted graph (not shown) derivative of graph given in Figure 5.13. The six measures

Table 5.2. Adjacency, average vertex degree, and connectedness of the nine functional groups of protein complexes in *Saccharomyces cerevisiae* (calculated from data of Gavin *et al.* [72])

Protein functional group	V	V ^a	A ^b	<a _i >	Conn	WA ^c	<w _i >	Wconn
RNA Metabolism	28	25	212	7.57	0.280	1124	40.14	1.487
Transcription/DNA Maintenance/Chromatin	55	44	468	8.50	0.158	1076	19.56	0.362
Protein Synthesis and Turnover	33	21	92	2.79	0.087	250	7.58	0.237
Membrane Biogenesis	20	11	40	2.00	0.106	44	2.20	0.116
Intermediate & Energy Metabolism	43	21	86	2.14	0.051	104	2.42	0.058
Protein RNA/Transport	12	6	12	1.00	0.091	20	1.67	0.152
Signaling	20	—	14	0.70	0.037	—	—	—
Cell Cycle	12	—	6	0.50	0.045	—	—	—
Cell Polarity & Structure	8	—	2	0.25	0.036	4	0.50	0.071

^aThe number of vertices in the giant component. No such component is available in the last three groups.

^bThe connectivity measures are calculated for the entire network, not for the giant component only.

^cThe calculations of the weighted indices is done with Eqs. (2.9) and (2.10), while those of the non-weighted indices by Eqs. (2.2) and (2.4).

Table 5.3. Complexity measures of six functional groups of protein complexes in *Saccharomyces cerevisiae* (calculated from data of Gavin *et al.* [72]): normalized second- and third-order subgraph count, first- and second-order overall connectivity, information on the vertex degree distribution, and A/D complexity index.

Protein functional group ^a	² SC	³ SC	¹ OC	² OC	I _{v,d,n}	A/D
RNA Metabolism	7.396	27.472	7.868	33.843	0.627	1.083
Transcription/DNA	0.605	0.650	0.631	0.729	0.522	0.289
Maintenance/Chromatin						
Protein Synthesis and Turnover	0.675	0.591	0.844	1.100	0.546	0.268
Membrane Biogenesis	0.200	0.095	0.224	0.115	0.422	0.216
Intermediate & Energy Metabolism	0.107	0.043	0.117	0.055	0.421	0.112
Protein RNA/Transport	0.517	0.312	0.640	0.448	0.477	0.385

^aThe functional groups of protein complexes involved in signaling, cell cycle, and cell polarity & structure are omitted, because they lack a giant component. The calculations are performed by the Grafrman software [80], making also use of Eqs. (3.9), (3.11), (3.19), and (3.20).

included: the two normalized subgraph count descriptors, 2SC and 3SC , the two normalized overall connectivity indices, 1OC and 2OC , the normalized information index for the vertex degree distribution, $I_{vd,n}$, and the newly developed A/D index, order the protein functional groups in a similar manner. They all single-out the functional group of protein complexes involved in the RNA metabolism as the most complex one, the next two places being occupied alternatively by the group controlling transcription, DNA maintenance, and chromatin structure, and the one of protein synthesis. The A/D index reproduces with a single exception the same ordering, and thus demonstrated its potential as complexity measure. It should be mentioned, that all our calculations were performed with data [72] that comprise about a quarter of all yeast proteins. Accounting for all protein complexes will indeed change the complexity measures values. One may anticipate that the availability of the complete set of data will enable the complexity estimates of performance stability of the biological functions related to cell cycle, cell polarity and structure, and signaling. One may also expect the major conclusion about the three groups of biological function that are best protected against any kind of damage to be confirmed by such more complete analysis.

6.2. Food webs

Food webs are presented by directed graphs, because the interaction between the species is in the great majority of cases unidirectional (the prey cannot eat the predator). Other examples of directed networks are gene regulatory networks and cellular signal transduction networks. It has been shown [64, 81, 82] that the more complex directed networks have a specific structure. It includes in- and out-components, a strongly connected component and a tube (Figure 5.14). The nodes in the *strongly connected component* are accessible to each other. These nodes have also incoming edges (arcs) originating from the *out-component*, and outgoing arcs directed to the *in-component*. Vertices from the in-component can also be directly connected to vertices of the out-component thus forming a *tube*.

As shown in the St. Martin island wood web [83], analyzed below (Figure 5.15), this specific hierarchical structure of directed networks is not always possible. The web incorporates 42 trophic species with a total of 205 interactions. The network of this ecological system is rather complex. Nevertheless, it does not have even a triplet of mutually accessible

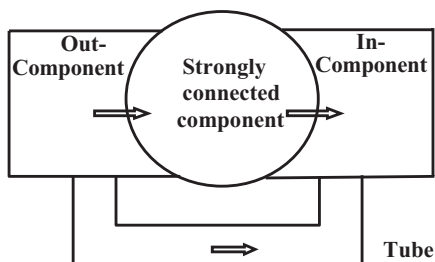


Figure 5.14. A typical structure of a complex directed graph.

vertices, which to form a strongly connected component. What appears as more essential and always preserved in such networks is their hierarchical structure, based on the principle of downstream interactions. In Figure 5.15, the St. Martin island wood web is presented schematically by two different directed graphs. The first graph is composed of six ordered layers A to F. The species of each layer can eat all downstream species, and in a few cases another species of the same layer. This graph shows explicitly only the interactions between the pairs of neighboring layers. The total amount of interactions between all pairs of layers is shown as edge weights in the second graph in Figure 5.15, the vertices in which depict the six layers of web species.

The connectivity of the St. Martin's island food web can be characterized by the values of the average vertex degree, $a_i = 4.88$, and that of connectedness (connectance) $= 0.119$. (Both values are just half of the corresponding values for the parent undirected graph.) The normalized 2SC and 3SC complexity indices are equal to 0.0673 and 0.0193, respectively. These values are the same for the directed graph and its undirected parent graph. The first three overall connectivity indices, 1OC , 2OC , and 3OC , are calculated in separate in- and out-terms (in: 0.0387, 0.0119, and 0.0037, and out: 0.0328, 0.0093, and 0.0028, respectively). The sums of the pairs of in- and out-terms, are equal to the corresponding parent undirected graph values. Therefore, the calculation of the in- and out-terms makes sense mainly when comparing different directed graphs DG_i originating from the same parent undirected graph G . In the case of DGs obtained from different Gs , one may use for approximate estimates the complexity measures as calculated for the corresponding parent graphs. The normalized information index on the vertex degree distribution also correctly reproduces the lower

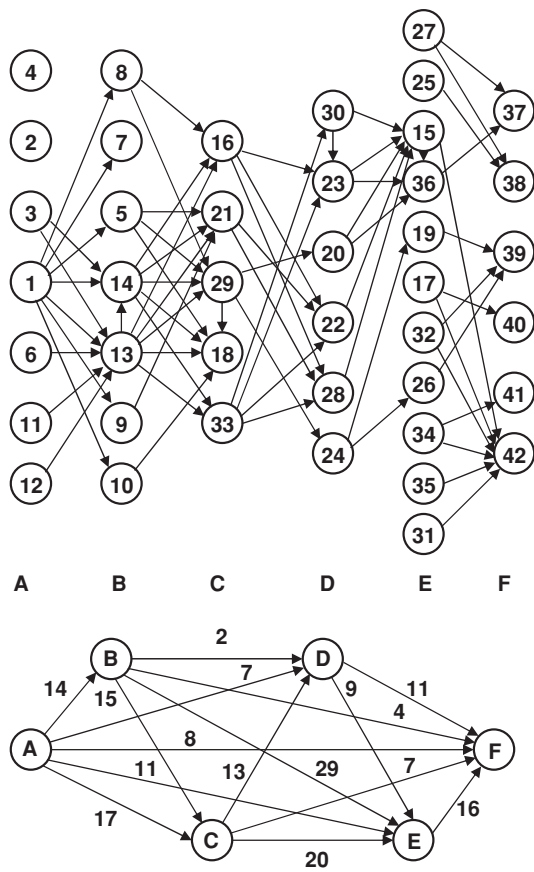


Figure 5.15. The connectivity of the StMartin island food web [83] is illustrated in two directed graphs formed by the hierarchically ordered layers A to F. The trophic species of each layer (numbered after [83]) can eat only downstream species and, in few cases, species of their own layer. The connectivity shown explicitly in the upper (unweighted) graph is that between the pairs of neighboring layers only. The edges of the lower (weighted) graph show the total number of interactions between all pairs of layers. The calculations of the complexity measures of the St. Martin's food web, however, are made proceeding from the entire directed graph with its 42 species and 205 directed interactions.

complexity of the directed graph relative to that of the parent undirected graph ($I_{vd,n}(out) = 0.367$, $I_{vd,n}(in) = 0.388$, $I_{vd,n}(G) = 0.401$).

There is no such correspondence between the distance measures of directed and parent undirected graphs, due to the lack of paths between some pairs of vertices in the *DGs*. Thus, while the undirected St. Martin graph has 1722 vertex-vertex distances, in the directed graph they are only 446 (205×1 , 209×2 , 32×3). The total distance calculated from these is 719 vs. 3308 in the undirected graph. Comparing the average distances of the two graphs would be misleading, because it would show the vertices of directed graph to be closer to each other than they are in the undirected graph (1.61 vs. 1.92). The things come back to normal after calculating the accessibility of the *DG* vertices (Eq. 5.1), $Acc = 0.259$, wherefrom Eq. (5.2) produces the more realistic value of $\langle d(DG) \rangle = 6.22 > 1.92$. More realistic estimate of the directed graph connectedness may also be obtained by Eq. (5.3), accounting for the limited vertex accessibility: $AConn(DG) = 0.031 < Conn(DG) = 0.119$, the latter value being unrealistically close to that of the undirected graph connectedness (0.238). Similar correction might be made for the *A/D* complexity index introduced in Section 4.1. This index shows a pattern of continuous increase with the increase in the network complexity. However, the value calculated for the directed graph, $A/D(DG) = 205/719 = 0.258$ is larger than that of the undirected graph, $A/D(G) = 410/3308 = 0.124$. The higher complexity of the undirected graph can be correctly assessed by adjusting the *A/D* index by multiplying it by the accessibility index ($0.258 \times 0.259 = 0.067 < 0.124$).

The different complexity indices order the food web in a similar manner (Figure 5.16). The connectedness index cannot distinguish two pairs of food webs (St. Martin Island/Lake Little Rock, $Conn = 0.119$, and Skipwith Pond/Coachella Valley, $Conn = 0.328/0.323$), whereas the latter are well discerned by the subgraph count and overall connectivity indices. Conversely, 2OC and 3OC cannot well discriminate Ythan Estuary and Canton Creek food webs.

Many studies have shown that a higher connectivity and complexity means a higher network stability [84,85]. One may thus expect the Skipwith Pond and Coachella Valley food webs to be very stable to attacks and environmental changes. As recently shown [45] the Skipwith Pond ecosystem could survive even the elimination of half of its best connected trophic species in the food web. The least complex webs examined—those

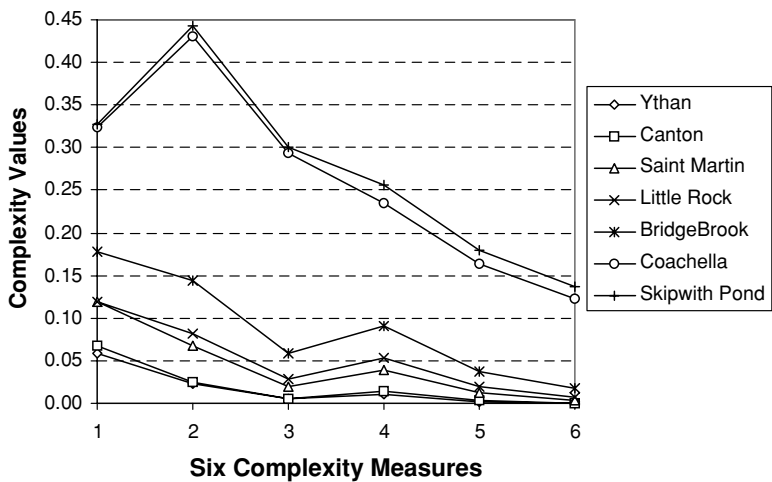


Figure 5.16. Complexity comparison of seven food webs (data from Dunne, Williams, and Martinez [83]) show the Skipwithpond and the Coachella Valley food webs to be the most complex ones, and the Canton Creek and Ythan Estuary to be the least complex ones. Complexity measures 1 to 6 correspond to connectedness (connectance), second- and third-order subgraph count, and first-, second-, and third-order overall connectivity [24-29].

of Ythan Estuary and Canton Creek—may be expected to be more vulnerable. To verify this conclusion, we modeled the specific attack on this web by subsequently eliminating its highest-degree vertices. It was found that after eliminating the first 13 such vertices, which corresponds to a 2-fold decrease in the web connectedness and to a 12-fold decrease in the web complexity as described by the 2SC and 1OC indices, the network splits into a large and a small component (Figure 5.17).

7. Overview

In this chapter, we reviewed some of the complexity measures, which were shown in previous publications to be appropriate for assessments of network complexity. A clear distinction was made between the two types of complexity: the compositional and the structural (topological) ones. Four topological complexity measures were presented in detail: the information on the vertex degree distribution, the subgraph count, the overall connectivity, and the walk count. The last three were presented as ordered sequences

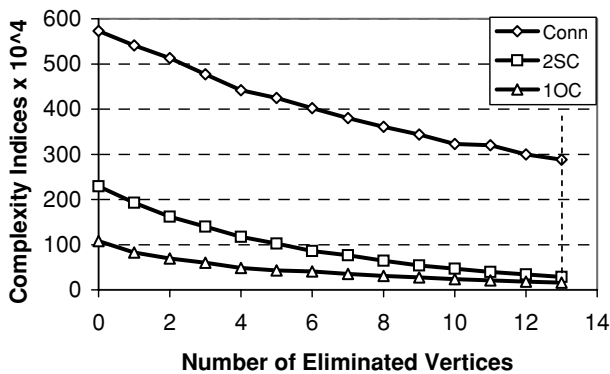


Figure 5.17. Stability analysis of the Ythan Estuary Food Web. The web splits into two pieces after eliminating the 13 highest connected vertices. The complexity measures used are the connectedness, the second-order subgraph count, and the first order overall connectivity.

of terms corresponding to subgraphs with increasing number of edges. Equations were derived for the first several orders of each of the complexity descriptors, which will facilitate their application to large scale networks. In addition, each of these measures was presented in three versions: total (or overall), average, and normalized (within the 0 to 1 range) ones. Two new complexity indices were proposed based on the combined use of the adjacency and distance matrix of the network. These indices unite the intuitive ideas of structural complexity resulting from high connectivity and small vertex separation (the “small world” concept). Important corrections were introduced to the way the total distance and the connectedness of directed graphs are calculated, by accounting for the mutual accessibility of network vertices. The mathematical tools introduced were illustrated with numerous examples, including protein-protein interaction networks and food webs. The authors anticipate a wider use of the presented complexity measures for the characterization of network topology, which usually does not go beyond connectedness (connectance), cluster coefficients, and graph radius.

Despite of the rapid development of complexity theory during the last 20 years, one can still face questions like: “Can we measure complexity, and, if we can, why?” We hope that this chapter answers explicitly the first question. As for the second one, we would like to remind the words of Lord Kelvin, said 150 years ago: “One cannot describe the Laws of Nature unless he uses numbers.” Are there laws of nature related to complexity

of systems? Up to very recently, there was no clear idea how to define complexity as a universal property of systems in nature and technology. The situation changed dramatically after Barabási [65] proposed in 1999 to consider the nonrandom dynamic networks as a universal language to describe complexity and evolution of systems. Life sciences have found in cellular networks (protein, gene, and metabolic ones) their long searched tool to describe the work of the biological machine as a whole. It is believed that the next 10-15 years will be the most important ones in the history of biology and medicine. The theory of network complexity will play an important role during this exciting time.

Acknowledgement

The authors are indebted to Drs. G. Rücker and C. Rücker (Bayreuth) for the use of their computer programs SUBGRATCAU and MOR5AU, and to Dr. J. A. Dunne (Santa Fe) for providing the food webs data. D. Bonchev was supported by NIH grant No. 5-22405.

References

1. C. Shannon and W. Weaver, *Mathematical Theory of Communications*, University of Illinois Press, Urbana, IL (1949).
2. H. Kastler (ed.), *Essays on the Use of Information Theory in Biology*, University of Illinois Press, Urbana, IL (1953).
3. H. Linshitz, The Information Content of a Bacterial Cell, in *Essays on the Use of Information Theory in Biology*, H. Kastler (ed.), University of Illinois Press, Urbana, IL (1953).
4. N. Rashevsky, *Bull. Math. Biophys.* **17**, 229-235 (1955).
5. E. Trucco, *Bull. Math. Biophys.* **18**, 129-135 (1956).
6. A. Mowshowitz, *Bull. Math. Biophys.* **30**, 175-204 (1968).
7. D. Minoli, *Atti. Acad. Naz. Lincei Rend.* **59**, 651-661 (1976).
8. A. N. Kolmogorov, *Problem'i Peredachi Informatsii* (Russ.) **1**, 1-7 (1965).
9. D. Bonchev, *Bulg. Chem. Commun.* **28**, 567-582 (1995).
10. D. Bonchev, D. Kamenski, and V. Kamenska, *Bull. Math. Biol.* **38**, 119-133 (1976).
11. D. Bonchev, and N. Trinajstić, *J. Chem. Phys.* **67**, 4517-4533 (1977).
12. D. Bonchev and V. Kamenska, *Croat. Chem. Acta* **51**, 19-27 (1978).
13. D. Bonchev, *MATCH - Commun. Math. Comput. Chem.* **7**, 65-113 (1979).
14. D. Bonchev, O. Mekenyan, and N. Trinajstić, *J. Comput. Chem.* **2**, 127-148 (1981).
15. D. Bonchev and N. Trinajstić, *Intern. J. Quantum Chem. Symp.* **16**, 463-480 (1982).
16. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Chichester, UK (1983).

17. D. Bonchev, Shannon's Information and Complexity in, *Mathematical Chemistry Series: Complexity in Chemistry*, Vol. 7, D. Bonchev and D. H. Rouvray (eds.), Taylor & Francis, London (2003) pp. 155-187.
18. S. H. Bertz, *J. Am. Chem. Soc.* **103**, 3599-3601 (1981).
19. S. H. Bertz, *J. Chem. Soc. Chem. Commun.* 209 (1981).
20. D. Bonchev, The Problems of Computing Molecular Complexity, in *Computational Chemical Graph Theory*, D. H. Rouvray (ed.), Nova Publications, New York (1990) pp. 34-67.
21. S. Nikolić, N. Trinajstić, M. Tolić, G. Rücker, and C. Rücker, On Molecular Complexity Indices, in *Mathematical Chemistry Series: Complexity in Chemistry*, Vol. 7, D. Bonchev and D. H. Rouvray (eds.), Taylor & Francis, London (2003) pp. 29-89.
22. S. H. Bertz, A. Mathematical Model of Molecular Complexity, in *Chemical Applications of Topology and Graph Theory*, R. B. King (ed.), Elsevier, Amsterdam (1983) pp. 206-221.
23. D. Bonchev and O. E. Polansky, On the Topological Complexity of Chemical Systems, in *Graph Theory and Topology in Chemistry*, R. B. King and D. H. Rouvray (eds.), Elsevier, Amsterdam (1987) pp.126-158.
24. D. Bonchev and W. A. Seitz, The Concept of Complexity in Chemistry. In: *Concepts in Chemistry: A Contemporary Challenge*, D. H. Rouvray (ed.), Wiley, New York (1997) pp. 353-381.
25. D. Bonchev, *SAR QSAR Environ. Res.* **7**, 23-43 (1997).
26. S. H. Bertz and T. J. Sommer, *Chem. Commun.* 2409-2410 (1997).
27. S. H. Bertz and W. F. Wright, *Graph Theory Notes New York Acad. Sci.* **35**, 32-48 (1998).
28. D. Bonchev, Overall Connectivity and Molecular Complexity. in *Topological Indices and Related Descriptors*, J. Devillers and A. T. Balaban (eds.), Gordon and Breach, Reading, UK (1999) pp. 361-401.
29. D. Bonchev, *J. Chem. Inf. Comput. Sci.* **40**, 934-941 (2000).
30. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **40**, 99-106 (2000).
31. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **41**, 1457-1462 (2001).
32. D. Bonchev, O. N. Temkin and D. Kamenski, *React. Kinet. Catal. Lett.* **15**, 119-124 (1980).
33. D. Bonchev, D. Kamenski, and O. N. Temkin, *J. Math. Chem.* **1**, 345-388 (1987).
34. K. Gordeeva, D. Bonchev, D. Kamenski, and O. N. Temkin, *J. Chem. Inf. Comput. Sci.* **34**, 244-247 (1994).
35. O. N. Temkin, A. V. Zeigarnik, and D. Bonchev, *Chemical Reaction Networks. A Graph Theoretical Approach*. CRC Press, Boca Raton, FL (1996).
36. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA (1969).
37. F. Harary, R. Z. Norman, and D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, Wiley, New York (1965).
38. N. Trinajstić, *Chemical Graph Theory*, 2nd ed., CRC Press, Boca Raton, FL (1992).

39. H. L. Morgan, *J. Chem. Docum.* **5**, 107-113 (1965).
40. D. Bonchev, A. T. Balaban and O. Mekenyan, *J. Chem. Inf. Comput. Sci.* **20**, 106-113 (1980).
41. D. Bonchev, *Theochem* **185**, 155-168 (1989).
42. D. Bonchev, O. Mekenyan and AT Balaban, *J. Chem. Inf. Comput. Sci.* **29**, 91-97 (1989).
43. M. E. J. Neuman, S. H. Strogatz, and D. J. Watts, *Random Graphs With Arbitrary Degree Distribution and Their Applications*, Santa Fe Institute (2000) Working Paper 00-07-042.
44. M. Gell-Mann, *The Quark and the Jaguar*, Freeman, New York (1994) p.31.
45. D. Bonchev, *SAR QSAR Envir. Sci.* **14**, 199-214 (2003).
46. D. Bonchev, Complexity of Protein-Protein Interaction Networks, Complexes and Pathways, in *Handbook of Proteomics Methods*, M. Conn (ed.), Humana, New York (2003) pp. 451-462.
47. D. Bonchev, *Chem. and Biodiversity*, **1**, 312-332 (2004).
48. D. Bonchev and D. H. Rouvray (eds.), *Mathematical Chemistry Series: Complexity in Chemistry*, Vol. 7, Taylor & Francis, London (2003).
49. S. Nicolíć, I. M. Tolić, N. Trinajstić, and I. Baučić, *Croat. Chem. Acta* **73**, 909-921 (2000).
50. M. Randić and D. Plavšić, *Croat. Chem. Acta* **75**, 107-116 (2002).
51. J. R. Platt, *J. Phys. Chem.* **56**, 328-336 (1952).
52. D. Bonchev, *Croat. Chem. Acta* **77**, 167-173 (2004).
53. M. Gordon and G. R. Scantlebury, *Trans. Faraday Soc.* **60**, 604-621 (1964).
54. S. H. Bertz and W. C. Herndon, The Similarity of Graphs and Molecules, in *Artificial Intelligence Applications to Chemistry*, T. H. Pierce and B. A. Hohne (eds.), ACS, Washington, DC (1986), pp.169-175.
55. G. Rücker and C. Rücker, *MATCH—Commun. Math. Comput. Chem.* **41**, 145-149 (2000).
56. D. Bonchev and N. Trinajstić, *SAR QSAR Environ. Res.* **12**, 213-235 (2001).
57. D. Bonchev, *J. Mol. Graphics Model.* **20**, 55-65 (2001).
58. H. Wiener, *J. Am. Chem. Soc.* **69**, 17-20 (1947).
59. H. Wiener, *J. Phys. Chem.* **52**, 1082-1089 (1948).
60. D. Bonchev, *J. Chem. Inf. Comput. Sci.* **41**, 582-592 (2001).
61. I. Gutman, B. Rušćić, N. Trinajstić, and C. W. Wilcox, Jr, *J. Chem. Phys.* **62**, 3399-3405 (1975).
62. S. Nicolíć, G. Kovacevic, A. Milicevic, and N. Trinajstić, *Croat. Chem. Acta* **76**, 113-124(2003).
63. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440-442 (1998).
64. A. L. Barabási, *Linked. The New Science of Networks*. Perseus, Cambridge, MA, (2002).
65. A. L. Barabási and R. Albert, *Science* **286**, 509-512 (1999).
66. S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079-1187 (2002).

67. <http://www.santafe.edu/sfi/publications/working-papers.html>.
68. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sasaki, *Proc. Natl. Acad. Sci. USA* **98**, 4569-4574 (2001).
69. G. Weng, U. S. Bhala, and R. Yyengar, *Science* **284**, 92-96 (1999).
70. A. Wagner, *Mol. Biol. Evol.* **18**, 1283-1292 (2001).
71. L. Giot, *et al.*, *Science* **302**, 1727-1736 (2003).
72. A. C. Gavin, *et al.*, *Nature* **415**, 141-147 (2002).
73. T. I. Lee, *et al.*, *Science*, **298**, 799-804 (2002).
74. N. Friedman, *Science*, **303**, 799-805 (2004).
75. A. H. Y. Tong, *et al.*, *Science*, **303**, 606-813 (2004).
76. H. Jeong, B. Tombor, Z. Albert, and A. L. Barabási, *Nature* **407**, 651-654 (2000).
77. A. Wagner and D. A. Fell, *Proc. Roy. Soc. London B* **268**, 1803-1810 (2001).
78. H. Ma and A. P. Zeng, *Bioinformatics* **19**, 270-277 (2003).
79. C. Koch and G. Laurent, *Science* **284**, 96-98 (1999).
80. S. Karabunarliev and D. Bonchev, Grafman software package, unpublished.
81. S. N. Dorogovtzev, J. F. F. Mendes, and A. N. Samukhin, *Giant Strongly Connected Component of Directed Graphs*, arXiv I cond-mat/0103629 v1 Mar 2001.
82. S. H. Yook, H. Jeong and A. L. Barabási, *Proc. Natl. Acad. Sci.* **99**, 13382-13386, (2002).
83. J. A. Dunne, R. J. Williams and N. D. Martinez, *Networks Topology and Biodiversity Loss in Food Webs: Robustness Increases With Connectance*, Santa Fe Institute Working Paper 02-03-013 (2002).
84. R. Albert, H. Jeong, and A. L. Barabási, *Nature* **406**, 378-382 (2000).
85. S. Maslov and K. Sneppen, *Science* **296**, 309-313 (2002).

Chapter 6

CELLULAR AUTOMATA MODELS OF COMPLEX BIOCHEMICAL SYSTEMS

Lemont B. Kier and Tarynn M. Witten

Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA
23284

1. Reality, Systems, and Models

1.1. Introduction

The role of a scientist is to study nature and to attempt to unlock her secrets. In order to pursue this goal, a certain process is usually followed, normally starting with observations. The scientist observes some part of the natural world and attempts to find patterns in the behaviors observed. These patterns, when they are found in what may be a quite complicated set of events, are then called the laws of behavior for the particular part of nature that has been studied. However, the process does not stop at this point. Scientists are not content merely to observe nature and catalog patterns; they seek explanations for the patterns. The possible explanations, that scientists propose, take the form of hypotheses and theories in the form of models. The models serve as representations about how things work behind the scenes of appearance. One way to describe the modeling process is to express it as a pictorial algorithm or flow diagram shown in Figure 6.1.

This chapter is about one such type of model and how it can be used to understand some of the more complex patterns of chemistry and biology. Let us begin by attempting to understand some essential principles behind modeling/simulation. We will then examine how, in certain scenarios, the models/simulations “*take on a life of their own*,” that is, they move from being *complicated* to being *complex*.

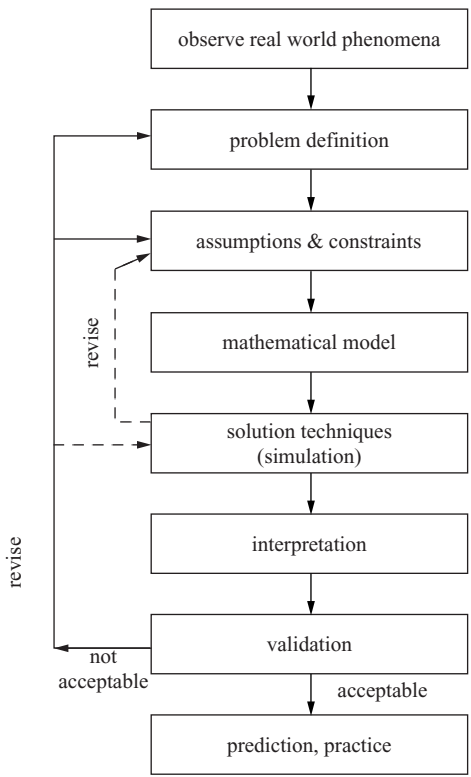


Figure 6.1. A flow diagram of the modeling process.

1.2. The “what” of modeling and simulation

A model is an observer/scientist’s attempt to represent, using a set of rules that they have deduced from observation and scientific deduction, the behavior of a system of interest. Consequently, a model is an abstraction of the whole system. By definition, due to its reduction, the model has access to fewer states than the original system. Hence, scientific interest lies in just a part of the whole system. Clearly, scientific logic dictates that the output of the model system should be consistent with the original system but only for a restricted set of inputs.

Many models in science take the form of mathematical relationships, equations connecting some property or set of properties with other parameters of the system. Some of these relationships are quite simple,

e.g., Newton's second law of motion, which says that force equals mass times acceleration, $F = ma$. Newton's gravitational law for the attractive force F between two masses m_1 and m_2 also takes a rather simple form,

$$F = \frac{Gm_1m_2}{r^2} \quad (1.1)$$

where r^2 is the square of the distance separating the masses and G is a constant that correctly dimensionalizes the units between the two sides of the equation. However, many mathematical relationships are much more complicated, and rely on the techniques of calculus, differential and partial differential equations, and abstract algebra to describe the rates of change of the quantities involved. Such an example would be the basic equation of quantum theory, the Schrödinger equation, which takes a more formidable form:

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \Psi = E\Psi \quad (1.2)$$

In chemical kinetics one finds linked sets of differential equations expressing the rates of change of the interacting species [1,2,3,4]. Overall, mathematical models have been exceedingly successful in depicting the broad outlines of an enormously diverse variety of phenomena in nature. Some scientists have even commented in surprise at how well mathematics works in describing nature. So successful have these mathematical models been that their use has spread from the hard sciences to areas as diverse as biology [5], medicine [6], economics [6], and the analysis of athletic performance [7]. In fact, many of the social and psychological sciences now use mathematical models to describe the behaviors of social systems [8,10], the spread of information in society, and the dynamics of psychological interaction [11].

In other cases models take a more pictorial form. In the early atomic models an atom was first pictured by J. J. Thomson as a "plum pudding", with negative electrons (the "plums") embedded in a spread-out positive charge (the pudding), and then later by Ernest Rutherford and Niels Bohr as a planetary system with a tiny positive core surrounded by circling electrons, a model called the "nuclear atom". Today, within quantum theory, the nuclear atom picture has been further transformed into one with a positive nucleus surrounded by a cloud of electron probabilities. In biology, the double-helix model of the structure of the genetic material DNA proposed

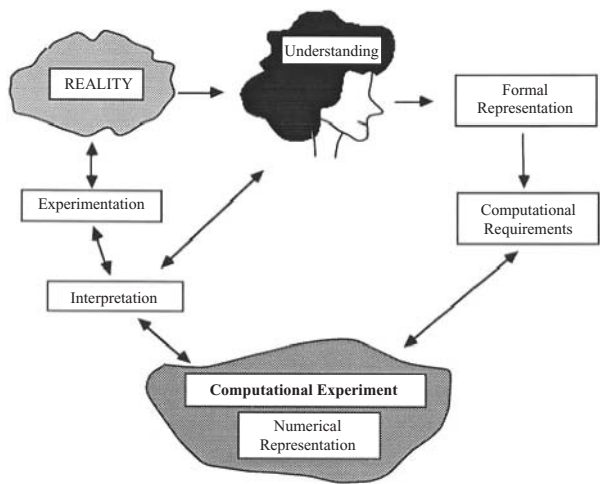


Figure 6.2. The modeling process including mental components.

by James Watson and Francis Crick led to an explosion of studies in the field of molecular genetics. Charles Darwin’s model of evolution by means of natural selection pictures species, composed of a collection of individuals with a variety of different traits, interacting with their environments. Individuals with some traits are better suited to survive and reproduce, thereby passing on these traits to their offspring. Over time new traits are introduced through mutations, environments gradually (or sometimes rapidly) change, and new forms develop from the old ones. The modern model of the human brain envisions regions devoted to different functions such as sight, motor movements, and higher thought processes. In geology, the tectonic plate model of the Earth pictures expansive continental plates moving gradually over the planet’s surface, generating earthquakes as they meet and slide over one another. And in psychology, the Freudian approach pictures human behavior as resulting from the actions of invisible components of the mind termed the id, the ego, and the superego. Thus, the modeling process could be pictorially represented as in Figure 6.2 [9].

The key feature of successful models is that they produce results consistent with the experimental observations. Successful models capture the essential features of the systems of interest. While it is not always true, scientists also hope that the newly created model will go beyond this simple reproduction to predict new features of the systems that may have

previously escaped notice. In this latter case the predictions provide an important means for testing the validity of the models. There are many different philosophical approaches to defining the art of model-building and its components. Consequently, at this point it is helpful to dissect models into their most significant parts, so that we can start from a common basis.

The system

Studies in chemistry, or any realm of science, commonly consist of a series of directed examinations of parts of nature's realm called systems. A *system* is an identifiable fragment of the world that is recognizable and that has attributes that one can identify in terms of form and/or function. We can give examples at any level of size and complexity, and in essentially any context. Indeed, a dog is a system at a pet show; whereas the human heart is a system to the cardiologist; a tumor cell is a system to the cancer specialist; a star or planet or galaxy is a system to an astronomer; a molecule, or a collection of molecules, is a system to a chemist; and a macromolecule in a cell is a system to a molecular biologist. A system is, then, whatever we choose to focus our attention upon for study and examination.

States of the system

A system is composed of parts that can be recognized and identified. As time goes by, a system under study may acquire different attributes as a result of changes among its parts, and over time its appearance or function may change. Moreover, as time goes by, our own technological capabilities may expand, thereby allowing us to identify parts of the system we would not have been able to previously identify. Each of the different stages through which the system passes in its evolution is called a state of the system. A dog grows old over time, passing through stages recognized in general terms as puppy, dog, and old dog. A heart may change its pattern of contractions, going from normal to tachycardia to ventricular fibrillation, each of which we categorize as different states of functioning. A solution of ethyl acetate in water may slowly decompose to mixtures of ethyl acetate, acetic acid and ethanol, through a sequence of states characterized by their different compositions. Water may start as a solid (ice), become a cool liquid, then a warmer liquid, and finally appear as a vapor at higher temperature, passing through these different stages as it is melted, heated, and vaporized. For the purposes of our discussion, we will refer to each

set of conditions as a *state* of the system under observation/study. It is the various states of the system upon which we focus attention when we study any system.

Observables

Our studies require us to analyze and describe the changes that occur in the systems we are interested in, as they evolve with time. To accomplish this analysis properly we need to record specific features that characterize what is occurring during the evolutionary process. The features assigned for this purpose are referred to as the *observables* of the study system. For example, we distinguish the puppy from the old dog by changes in its physical appearance and its behavior. The changes in a heart's rhythm are recorded on special charts monitoring electrical signals. The changes occurring in a solution of ethyl acetate in water can be characterized by changes in the solution's acidity, by spectroscopic readings or by detection of the odor of acetic acid. To be as precise as possible in a scientific investigation it is necessary to assign numerical values to the characteristics that distinguish one state of a system from another.

The state of a system is studied through detection and recording of its observables. To record an observable, we must "*probe*" the system with a "*measurement*" instrument of some type. This requires an "*interaction*," which we will discuss in the next section, as well as the existence of a device that is capable of recognizing the particular system observable as well as reporting back to the scientist/observer the value of that observable. It is extremely important to understand that observables of a system are intimately tied to the technological capability of the scientist. Thus, gene sequencing, common in today's technological arsenal, was not a probe available to scientists fifty years ago. Thus, if our observable was the gene sequence of an organism, the probe did not exist to provide the requisite information. Hence, the observable, while of scientific interest, was not accessible.

Interactions

There are two types of *interaction* with which we must deal. The first is the interaction of the scientist/observer with the system under study. The second is the interaction of the parts of the system with each other (both known and unknown).

The scientist interacts with the system in two ways; through setting the actual experiment up to be observed and through measurement probing of the system. Each of these interactions can obviously affect how the scientist sees the system and thereby subsequently affect the resultant measurements and through this, affect the scientific observations, conclusions and the modeling effort. For example, removing a wolf from the pack may help you to understand the isolated wolf's health status, but does not tell us anything about how the social hierarchy of the wolf pack affects the wolf's health. Thus, if you do not know that social hierarchy and dynamics is important, the single wolf experiment removed an interaction necessary for the study of the wolf and consequently affects the conclusions available to the scientist, thereby affecting the accuracy of the conclusions and any subsequent model-building exercise.

The second interaction, the scientist-system interaction, is more subtle. However, it must be mentioned. The act of actually inserting a probe into a system clearly affects the system. Thus, the scientist is forced to ask the question of whether or not the measurements being obtained are actually those of the isolated system of interest or of the system-probe complex. This question, while seemingly philosophical, has important ramifications in quantum-level measurements and in abstract theoretical biology [12,13].

The parts of a complex system naturally interact with one another, and the fascinating evolutionary dynamics of complex systems depends crucially upon the nature of these interactions. The interactions supply the driving forces for the changes that we observe in systems. In addition, we can change the behavior of a system by introducing new elements or ingredients. Intrusions of this kind produce new interactions, which in turn alter the system. By carefully choosing the added factors and interactions we, as scientists, can develop new patterns of observables that may be revealing. Interaction with your dog might include exercising to increase his running stamina, which in turn will lead to a new, improved set of health-related (state) indicators. Electrical stimulation of a fibrillating heart can introduce interactions that lead to the conversion of the heart from the fibrillating state to a normal, healthy state of performance. Heating the ethyl acetate solution will eventually accelerate the hydrolysis reaction and distill away the resulting ethanol, leaving a solution of acetic acid. The interactions introduced and the accompanying changes in a system's observables produce information about the nature of the system and its behavior under different

conditions. With enough observables we may be able to piece together a reasonable description, a model, of how the system operates.

1.3. Back to models

From a carefully selected list of experiments with a system we can evoke certain conclusions. The mosaic of information leads us to piece together a description of the system, what is going on inside it, the relationships among its states, and how these states change under different circumstances. In the case of our dog, the exercise tests may lead us to theorize that the dog is in good or poor health. With the heart, the electrical impulses that we record can reveal a pattern of changes (observables) that we theorize to belong to a healthy (or diseased) heart. By subjecting the solution of hydrolyzing chemicals to fractional distillation and chemical analysis we may theorize that we originally had a system of water and ethyl acetate.

We can arrive at our theories in two main ways. In the first, as illustrated above, we subject a system to experimental perturbations, tests, and intrusions, thereby leading to patterns of observables from which we may concoct a theory of the system's structure and function. An alternative approach, made possible by the dramatic advances that have occurred in the area of computer hardware in recent times, is to construct a computer model of the system and then to carry out simulations of its behavior under different conditions. The computer "experiments" can lead to observables that may be interpreted as though they were derived from interactions.

Simulations

It is important to recognize the different concepts conveyed by the terms "model" and "simulation", even though these terms are sometimes used interchangeably. As noted above, a model is a general construct in which the parts of a system and the interactions between these parts are identified. The model is necessarily simpler than the original system, although it may itself take on a rather complicated form. It consists of ingredients and proposals for their interactions.

Simulations are active imitations of real things, and there are generally two different types of simulations, with different aims. In one approach a simulation is merely designed to match a certain behavior, often in a very limited context. Thus a mechanical bird whistle may simulate a sound resembling that of a bird, and does so through a very different mechanism

than the real source of the sound. Such a simulation reveals little or nothing about the features of the original system, and is not intended to do so. Only the outcome, to some extent, matches reality. A hologram may look like a real object, but it is constructed from interfering light waves.

A second type of simulation is more ambitious. It attempts to mimic at least some of the key features of the system under study, with the intent of gaining insight into how the system operates. In the context of our modeling exercise, a simulation of this sort means letting our model “run.” It refers to the act of letting the parts of our model interact and seeing what happens. The results are sometimes very surprising and informative.

Why are modeling and simulation important?

Beginning in the late 1800’s, mathematicians began to realize that biology and ecology were sources of intriguing mathematical problems. The very complexity that made life difficult for experimental biologists intrigued mathematicians and led to the development of the field of Mathematical Biology. More recently, as computers became more cost-effective, simulation modeling became more widely used for incorporating the necessary biological complexity into the original, often over-simplified mathematical models.

The experimentalists felt that the theoretical analyses were deficient in a variety of areas. The models were far too simple to be useful in clinical or practical biological application. They lacked crucial biological and medical realism. Mathematical modelers balked at the demands for increased levels of biological complexity. The addition of the required biological reality, desired by the life scientists, often lead to alterations in the formulation of the mathematical models, alterations that made the models intractable to formal mathematical analysis.

With the advent of the new high performance computer technologies and the deluge of ‘omic’-data, biological and biomedical reality is finally within the grasp of the bio-modeler. Mathematical complexity is no longer as serious an issue, as new mathematical tools and techniques have grown at nearly the same speed as the development of computational technology [14].

The role of high performance computing and modeling in the sciences has been documented by numerous federal publications (NSF [15,16], for example). Many of the grand challenge bio-computational problems of the 1990’s still remain so. Some of these problems, such as multi-scale

simulations, and such grand challenge computational problems as linking heart and kidney simulations, which are now beginning to become addressable, were only pipedreams during the 1990's [17-19]. More recently, such models and their simulations are being termed "*in silico*" modeling and simulation.

Modeling and simulation provide the scientist with two very useful tools. The first of these is validation of the theoretical understanding and its model implementation. The second of these tools is that, the more complete the model, the more it provides an experimental laboratory for further research on the very system being modeled. Thus, "*in silico*" models can both validate current viewpoints/perspectives of the dynamical evolution of a system and can provide an environment in which the scientist can explore potential new theories and their consequences. It is this second aspect of models and their simulations that is of particular interest to us. Let us take a moment to address modeling in chemistry and molecular biology.

1.4. Models in chemistry and molecular biology

Chemistry and molecular biology, like other sciences, progresses through the use of models. They are the means by which we attempt to understand nature. In this chapter we are primarily concerned with models of complex systems, those whose behaviors result from the many interactions of a large number of ingredients. In this context two powerful approaches have been developed in recent years for chemical investigations: molecular dynamics and Monte Carlo calculations [20-25]. Both techniques have been made possible by the development of extremely powerful, modern, high-speed computers.

Both of these approaches rely, in most cases, on classical ideas that picture the atoms and molecules in the system interacting *via* ordinary electrical and steric forces. These interactions between the ingredients are expressed in terms of force fields, *i.e.*, sets of mathematical equations that describe the attractions and repulsions between the atomic charges, the forces needed to stretch or compress the chemical bonds, repulsions between the atoms due to their excluded volumes, etc. A variety of different force fields have been developed by different workers to represent the forces present in chemical systems, and although these differ in their details, they tend generally to include the same aspects of the molecular interactions. Some are directed more specifically at the forces important for, say, protein

structure, while others focus more on features important in liquids. With time more and more sophisticated force fields are continually being introduced to include additional aspects of the inter-atomic interactions, *e.g.*, polarizations of the atomic charge clouds and more subtle influences associated with quantum chemical effects. Naturally, inclusion of these additional features requires greater computational effort, so that a compromise between sophistication and practicality is required.

The molecular dynamics approach has been called a brute-force solution of Newton's equations of motion [20]. One normally starts a simulation using some assumed configuration of the system components, for example an X-ray diffraction structure obtained for a protein in crystalline form or some arrangement of liquid molecules enclosed in a box. In the protein case one might next introduce solvent molecules to surround the protein. One then allows the system, protein-in-solvent or liquid sample, to evolve in time as governed by the interactions of the force field. As this happens one observes the different configurations of the species that appear and disappear. Periodic boundary conditions are usually applied such that molecules leaving the box on the right side appear on the left side; those leaving at the top appear at the bottom, and so forth. The system's evolution occurs *via* time steps (iterations) that are normally taken to be very short, *e.g.*, 0.5–2.0 femtoseconds (fs, 10^{-15} seconds), so that Newton's second law of motion

$$F = ma = m \frac{dv}{dt} \quad (1.3)$$

can be assumed to hold in a nearly linear form. The evolution of the system is followed over a very great number of time steps, often more than a million, and averages for the features of interest of the system are determined over this time frame. Because the calculation of the large number of interactions present in such a system is very computationally demanding, the simulations take far longer than the actual time scale of the molecular events. Indeed, at present most research-level simulations of this type cover at best only a few tens of nanoseconds of real time. (Note that 10^6 steps of 1 fs duration equal one nanosecond, 10^{-9} s.) Whereas such a timeframe is sufficient to examine many phenomena of chemical and biochemical interest, other phenomena, which occur over longer time scales, are not as conveniently studied using this approach.

The Monte Carlo method for molecular simulations takes a rather different approach from that of the molecular dynamics method [24–26]. Rather

than watching the system evolve under the influence of the force field, as done in molecular dynamics, a very large number of possible configurations of the system are sampled by moving the ingredients by random amounts in each step. New configurations are evaluated according to their energies, so that those lowering the energy of the system are accepted whereas those raising the system energy are conditionally “weighted”, or proportionately accepted, according to their potential energies. The weighting is normally taken to have the form of the Boltzmann distribution, *i.e.*, to be proportional to $e^{-\Delta V/kT}$, where ΔV is the potential energy change, k is Boltzmann’s constant, and T is the absolute temperature. From statistical analysis of a large, weighted sample (ensemble) of such configurations one can ascertain many of the important thermodynamic and structural features of the system. A typical sample size employed for this purpose might encompass between one and ten million configurations.

Both the molecular dynamics and the Monte Carlo approaches have great strengths and often lead to quite similar results for the properties of the systems investigated. However, these methods depend on rather elaborate models of the molecular interactions. As a result, as noted above, both methods are highly computationally demanding, and research-level calculations are normally run on supercomputers, clusters, or other large systems. In the next section we shall introduce an alternative approach that greatly simplifies the view of the molecular system, and that, in turn, significantly reduces the computational demand, so that ordinary personal computers suffice for calculations and elongated time frames can be investigated. The elaborate force fields are replaced by simple, heuristic rules. This simplified approach employs another alternative modeling approach using *cellular automata*. However, before we begin this discussion, we must first address the general subject of complexity.

2. General Principles of Complexity

2.1. Defining complexity: complicated vs. complex

Up to this point we have been discussing “*complicated*” systems whereby complicated, we mean that they may be organized in very intricate ways, but they exhibit no properties that are not already programmed into the system. We may summarize this by saying that complicated systems are no more than the sum of their parts. Moreover, should we be able to isolate all of the parts and provide all possible inputs, we would, in theory, know

everything that there is to know about the system. Complicated systems also have the property that one key defect can bring the entire system to its knees. Thus, in order to make sure that such a problem does not occur, the system must have built-in redundancy. Redundancy is necessary because complicated systems do not adapt. A familiar example here is household plumbing. This is a relatively complicated system (to me) where there are many cut-off valves that can be used to deal with a leak. The leak does not solve itself.

There are systems, however, where the “whole is greater than the sum of the parts.” Systems of this type have the property that decomposing the system and analyzing the pieces does not necessarily give clues as to the behavior of the whole system.” We call such systems, “*complex*” systems. These are systems that display properties called “*emergence*,” “*adaptation*,” and “*self-organization*.” Systems that fall under the rubric of complex systems include molecules, metabolic networks, signaling pathways, ecosystems, the world-wide-web, and even the propagation of HIV infections. Ideas about complex systems are making their way into many fields including the social sciences and anthropology, political science and finance, ecology and biology, and medicine. Let us consider a couple of familiar examples. Consider a quantity of hydrogen and oxygen molecules. Gas laws are obeyed; the system can be defined by the nature of both gases. We would call this a simple system.



If we now ignite the system producing a reaction leading to water, we now have a complex system where knowledge of H_2 and O_2 no longer tell us anything about the behavior and properties of H_2O . The properties of water have emerged from the proto-system of the two gases. The two gases have experience the dissolvence of their properties in this process, an event that will be described later. A second familiar example is the array of amino acids, twenty in nature, that are available for polymerization. When this process occurs, a large, new molecule, a protein, appears. The behavior and properties of the protein are not discernable from a simple list of the amino acids.



The spatial structure and functions of this protein are non-linear functions of the kinds and numbers of the amino acids and their order of linkage in

the protein. The amino acids have surrendered their individual properties and functions, blending these into the whole, the protein molecule.

In order to understand the distinction between complex and complicated systems, we need to make some definitions.

2.2. Defining complexity: agents, hierarchy, self-organization, emergence, and dissolution

Agents

The concept of an agent emerges out of the world of computer simulation. Agent-based models are computer-driven tools to study the intricate dynamics of complex adaptive systems. We use agent-based models because they offer unique advantages to studying complex systems. One of the most powerful of these advantages is the ability of such modeling systems to be used to study complex social systems, complex biological and biomedical systems, molecules, and even complex financial systems that we could not model using mathematical equations or which may be intractable mathematically. Agent-based approaches allow us to examine, not only the final outcome of a simulation, but the whole history of the system as the interactions proceed. Moreover, agent-based models allow us to examine the effects of different “*rules*” on a system.

An *agent* is the lowest level of the model or simulation. For example, if the environment is a checkerboard, then the agents are the checkers; if it is a chess board, then the agents are the chess pieces. Thus, agents act within their environment. However, this is an extremely broad definition of an agent. Let us look at the structure of agents in a closer fashion.

The agent exists within the environment; the agent interacts with the environment/other agents by performing actions on/within it, the environment/other agents may or may not respond to those actions with changes in state. Agents are assumed to have a repertoire of possible actions available to them [27]. These actions can change the state of other agents or the environment. Hence, if we were to look at a basic model of agents interacting within their environment, it would look as follows. The environment and the agents start in an initial configuration/set of states. The agent begins by choosing an action to perform on other agents or on the environment. As a result of this action, the environment/other agents can respond with a number of possible states. What is important to understand is that the outcome of the response is not predictable. On the basis of the response

received, the agent again chooses an action to perform. This process is repeated over and over again. Because the interaction is history dependent, there is a non-determinism within the system.

Agents do not act without some sort of rule-base. We build agents to carry out tasks for us. Depending upon what is being modeled, the rule-base can be simple interaction rules or it can be more sophisticated rules about achievement, maintenance, utility, or other performance rules. These provide rules about how the agents function in the environment. Checkers have rules about jumping, kinging, and movement; similarly, chess pieces have rules. While checkers consists of one type of agent (*homogeneous*), “the checker,” with a simple set of rules, chess is a “*multi-agent*” (*heterogeneous*) game having different agents with different interaction rules. Closer to home we recognize the rules, called valence, that proscribe the bonding patterns of atoms to form molecules. Additionally, agent interaction rules can be static (unchanging over the lifetime of the simulation) or dynamic. They can operate on multiple temporal and spatial scales (local and global). They can be direct, indirect, or even hierarchical. And, one can even assume a generalized form of inter-agent communication by allowing the agents to see the changes caused by other agents and to alter their operational rules in response to those observations. In the upcoming section on cellular automata, we will illustrate some of these concepts in more detail.

Hierarchy

The concept of *hierarchy* is intuitive [28,29] If we look at complex systems, we see that they are made up of what we might call “layers” of structure. The human body contains numerous examples of hierarchical structures. The excretory system contains numerous organs. Those organs contain numerous cells, those cells contain numerous subcellular metabolic and signaling systems, and those systems contain numerous atoms and molecules. At each level of “organization,” there are rules, functions, and dynamics that are being carried out. Similar arguments can be made about the cardiovascular system, the nervous system, the digestive system, etc. Even the brain can be subdivided in a hierarchical fashion. The brain is formed from the cerebrum, the cerebellum, and the brain stem. The cerebrum is divided into two hemispheres. Each hemisphere is divided into four lobes. Each lobe is further divided into smaller functional regions [30] And again, in each area of the brain, and at all levels, the “brain” system

is engaging in various hierarchically related behaviors. Other examples of hierarchies include ecosystems [31] and social systems [9].

What makes hierarchies interesting is not just that they exist, but also how they are structured and how the levels of the hierarchies are interconnected. Moreover, one can ask how hierarchies evolved into the particular forms that they currently display. Additionally, one can ask questions concerning how the functions of the network evolved as they did. These questions lead us to ask about relational aspects of a hierarchy/network, self-organizational aspects of the hierarchy/network, and emergent properties of such systems. Some of these questions are addressed, in greater detail, in the chapter by Don Mikulecky in this volume. However, what is important to understand is that hierarchies have both temporal and spatial organization and that these organizational forms create the pathways for patterns of behavior and, as we shall see in a moment, these patterns of behavior and structural forms are often not predictable; they are self-organizational and emergent. Some of the groundbreaking, original work in this field was done by Nicolas Rashevsky [32] and Robert Rosen [33-36] (who was Rashevsky's student). Their work involved examining biological systems from the standpoint of agents (although at that time they were not called that) and the relationships between the agents. Rosen's work was extended by Witten [37] (who was Rosen's student) in an effort to study the dynamic complexity of senescence in biological systems. A simple argument is as follows. It is certainly clear that every biological organism or system O is characterized by a collection P of relevant biological properties P_i which allows the observer to recognize our biological organism as a specific organism. That is, these properties allow us to distinguish between organisms. This collection of properties P_i is the set of biological properties representing the organism O . It can be the very "coarse" set of sensitivity (S) to stimuli, movement (M), ingestion (I), and digestion (D). Or, at a slightly less coarse (more detailed) level, we might consider the set of all hormones (H) in an organism, and the set of that organism's metabolic responses (R). It is also clear that many of our biological properties P_i will be related to each other in some way. For example, ingestion (I) must come before digestion (D). We may denote this ordering through the use of arrows as follows, $I \rightarrow D$. For those readers who are familiar with business management, an organizational chart is a perfect example of such an interrelated collection of properties. If we were to say that two

properties were related, but not indicate how, we would write I-D. We call such figures *graphs*. When the edges of the graph are directed, we term such graphs *directed graphs*. The elements at the intersection of two or more edges are called the nodes or vertices of the graph. Hence, biological hierarchies may, in some cases, be represented by directed graphs: or what we might call *dependency networks* [37]. In summary, we have seen that abstracting away the fine grain aspects of biological systems allows us to represent their complexity (hierarchical structure) by either directed or undirected graphs.

Why is such an approach important? First, it allows us to represent basic processes of the system without being involved in the details at microscopic levels. Such a representation is useful, particularly if the mathematical modeling becomes intractable to analysis. More important is the fact that using hierarchical representation of a system allows us to understand *relationships* between components in a way that is not amenable to traditional mathematical modeling and simulation techniques. It is not so much about what the boxes in the network actually do as about how they are interconnected and how that set of connections creates the possibility for various dynamical behaviors. This approach is currently undergoing a great resurgence with the new studies of genomic [38], metabolic [39,40], and other networks [41]. In fact, this approach is being used to study ecological systems such as food webs [42], human sexual contacts and linguistics [43], and even e-mail networks and telephone networks [44]. Biochemical systems can also be studied with these “*topological*” approaches. Seminal work in this area has been done by Bonchev and his collaborators [45,46]. These structural or topological approaches are generalizable across systems spanning vast orders of hierarchical magnitude. One could say that one of the major characteristics of complex systems is that there are common behaviors, a number of levels or scales that dynamically interact and have many components. A marvelous example of such a complex system can be found in dealing with hierarchical modularity in the bacterial *E. coli* [47]. In this paper, the authors demonstrate that the metabolic networks of 43 distinct organisms are organized into many small, highly connected, topologic modules that combine in a hierarchical manner into large, less cohesive units. It follows then that, within the biochemical pathways of metabolic networks, there is a large degree of hierarchical structure [48].

As a consequence of such topological organizational properties, networks generate properties that cannot be simply inferred from the behavior of the components (*emergence*). They develop unpredictable temporal behavior (*chaos*). And they develop the ability to organize themselves in ways that were not obvious from the component pieces (*self-organization*). Let us briefly address these three properties and then illustrate them by focusing on cellular automata models of biochemical systems.

Self-organization and emergence

Self-organization and emergence are two properties of complex systems that are very much intertwined. Like hierarchy, their meaning seems intuitive. And yet, there is far more to self-organization and emergence than can be easily reviewed, much less covered in this brief chapter. Let us begin with the concept of self-organization. Stuart Kauffman, one of complexity theory's greatest proponents spoke of self-organization in the following fashion, "*Self-organization* is matter's incessant attempts to organize itself into ever more complex structures, even in the face of the incessant forces of dissolution described by the second law of thermodynamics [49]. By means of a simple example, suppose we have the following set of letters, L, S, A, T, E. Moreover, suppose that each of them was written on a card that had magnets placed on its four edges. Obviously, just sitting there, the letters have no intrinsic value other than representing certain sounds; they exist as a collection of objects. If, however, we put them into a shoebox, shake the box, and let them magnetize to each other, we might obtain any number of letter combinations; LTS, ATE, TLSA, STLAE, STALE, etc. Again, intrinsically, these organized structures have no meaning. At the lowest hierarchy level, they represent new organized structures that occurred because we shook them around in a shoebox. Suppose, however, that we now give them context. That is, at a higher hierarchical level, that of language, these strings of letters acquire a new property; that of meaning. Meaning, through the self-organization process of being shaken around in the shoebox, becomes an emergent property of the system. It could not have been inferred from the simple lower-level collection of letters. Suddenly, some of the organized structures, like ATE and STALE, lose their sense as strings of valueless symbols and acquire this new property of being a word with meaning. Similarly, one can make the same argument by putting the magnetized words into the shoebox and creating strings of words. Some of the word strings will acquire meaning and will be called sentences such

as ELSA ATE STALE TEA. Perhaps, along with exogenous factors, a language evolves. Language, the emergent property of the interaction of the letters and the environment/culture, could never have been predicted from the properties of the magnetic letters themselves. Nor could it have been predicted from the strings of letters. Thus language itself could be viewed as an emergent property.

Well, at this point, you are wondering what this has to do with anything chemical. Let's take a look at some examples from the biological and chemical world. First we consider a very simple example. Atoms, the lowest hierarchical unit, have certain properties. These properties allow them to combine, in various ways, to form molecular structures. When this happens, the properties of the atoms are lost, in part, to the overall molecular properties. Another simple example is the laser. A solid state laser consists of a rod of material in which specific atoms are embedded. Each atom may be excited by energy from outside, leading it to the emission of light pulses. Mirrors at the end of the rod serve to select these pulses. If the pulses run axially down the rod, then they will be reflected several times and stay longer in the laser. Pulses that do not emit axially leave the laser. If the laser is pumped with low power, the rod will illuminate, but it will look more like a lamp. However, at a certain pumping power, the atoms will oscillate in phase and a single pulse of light will be emitted. Thus, the laser is an example of how macroscopic order emerges from self-organization. What is interesting about this particular example is that the order is not near equilibrium for the system. In fact, the laser beam is dissipative and far from thermal equilibrium.

Other examples of self-organization occur in the kinetics of the autocatalytic formation of sugars from formaldehyde (formol reaction) [50]. Radical new self-organizational behaviors have been discovered in numerous biochemical systems; enzyme reactions [51], glycolysis [52,53], and the Bray-Liebfafsky reaction of iodate and hydrogen peroxide [54] to name just a few. One of the most striking of these reactions and certainly one of the most famous is the Belousov-Zhabotinsky reaction [55]. What happens is that under certain non-equilibrium conditions, this system behaves with all sorts of unexpected and unpredictable behaviors. In terms of its reactants, the BZ-system is not an unusual one. A typical preparation consists of cerium sulfate, malonic acid, and potassium bromate, all dissolved in sulfuric acid. It is easy to follow the pattern formation because an excess of cerium Ce^{4+} ions gives a pale yellow color to the solution, where as if there

is an excess of Ce^{3+} ions, the solution is colorless. Depending upon the initial mixing conditions, ionic potential traces show sustained oscillations, damped oscillations, and chaotic oscillations. When viewed spatially, the reaction displaces spiral waves, some of which have multi-arm spirals.

Cellular and subcellular biochemical signaling pathways are also extremely complex. They allow the cell to receive, process, and to respond to information. Frequently, components of different pathways interact and these interactions result in signaling networks. Under various conditions, such networks exhibit emergent properties that are; they exhibit properties that could not have been inferred from the behaviors of the parts. Such properties include integration of signals across multiple time-scales, generation of distinct outputs depending upon input strength and duration, and self-sustaining feedback loops. Moreover, the feedback can result in bistable behavior with discrete steady-state activities not available to any of the component pieces. One of the consequences of emergent properties is that it raises the possibility that information for learned behavior of biological systems may be stored within intracellular biochemical reactions that comprise signaling pathways [56].

Emergence

A corollary to emergence is the loss of identity, properties, and attributes (called property space) of the agents as they progressively self-organize to form complex systems at a higher hierarchical level. Testa and Kier have addressed this issue where they have referred this reciprocal event as dissolvence [57,58]. It is the reduction in the number of probable states of agents as they engage each other in the synergy with fellow agents. This is a partial loss; they do not disappear but are dissolved into the higher system. As hydrogen and oxygen are consumed in a reaction to form water, these atoms lose their identity as gases with free movement and become joined with each other to change state and to become ensnared in a fixed relationship. To quote H. G. Wells and J. S. Huxley [59]: *He escapes from his ego by this merger and acquires an impersonal immortality in the association; his identity **dissolving** into greater identity.*

The next step

In the preceding discussion, we have seen how complex behaviors can emerge from the combination of simple systems. Moreover, we have seen

how these behaviors are not predictable from the pieces that compose the system. This raises the following question. How does one model such behaviors? In the next section, we will address one modeling approach to handling systems that might display self-organization and emergence phenomena, that of cellular automata.

3. Modeling Emergence in Complex Biosystems

3.1. Cellular automata

Cellular automata were first proposed by the mathematician Stanislaw Ulam and the mathematical physicist John von Neumann a half century ago [60-62] although related ideas were put forth earlier by the German engineer Konrad Zuse [63]. Von Neumann's interest was in the construction of self-reproducing automata. His idea was to construct a series of mechanical devices or automata that would gather and integrate the ingredients that could reproduce themselves. A suggestion by Ulam [61] led him to consider grids with moving ingredients, operating with rules. The first such system proposed by von Neumann was made up of square cells in a matrix, each with a state, operating with a set of rules in a two-dimensional grid. With the development of modern digital computers it became increasingly clear that these fairly abstract ideas could in fact be usefully applied to the examination of real physical systems [64-67]. As described by Wolfram [68] cellular automata have five fundamental defining characteristics:

- They consist of a discrete lattice of cells
- They evolve in discrete time steps
- Each site takes on a finite number of possible values
- The value of each site evolves according to the same rules
- The rules for the evolution of a site depend only on a local neighborhood of sites around it

As we shall see, the fourth characteristic can include probabilistic as well as deterministic rules. An important feature sometimes observed in the evolution of these computational systems is the development of unanticipated patterns of ordered dynamical behavior, or emergent properties to be used to drive further experimental inquiry.

Cellular automata are one of several approaches to the modeling of complex dynamic systems. It is a model because it is used as an abstraction

of a system in which a portion has been selected for study. The principle features are the modeling of state changes and/or the movement of parts of a system. Such a simple model would be expected to have a very wide range of applications in nature. Indeed, there are many studies reported in the literature such as music, ants traffic, cities and so forth.

The results of a CA model are new sets of states of the ingredients called the configuration of the system. This configuration arises from many changes and encounters among the ingredients of the CA. These changes may occur over a very long period of “time” in the model. The ability of a computer to carry out many changes simulating a long time is a huge advantage of the machine. Before the computer, it would have taken unfathomable amounts of individual calculations over a vast amount of real time. The CA then, is a platform on which many changes can take place, data collected and reported.

3.2. The general structure

The simulation of a dynamic system using cellular automata requires several parts that make up the process. The cell is the basic model of each ingredient, molecule or whatever constitutes the system. These cells may have several shapes as part of the matrix or grid of cells. The grid containing these cells may have boundaries or be part of a topological object that eliminates boundaries. The cells may have rules that apply to all of the edges or there may be different rules for each edge. This latter plan may impart more detail to the model, as needed for a more detailed study. This is a grid with ingredients, A and B occupying the cells shown in Figure 6.3.

The Cells

Cellular automata have been designed for one, two or three-dimensional models. The most commonly used is the two-dimensional grid. The cells may be triangles, squares, hexagons or other shapes in the two-dimensional grid. The square cell has been the one most widely used over the past 40 years. Each cell in the grid is endowed with a primary state, i.e., whether it is empty or occupied with a particle, object, molecule or whatever the system requires to study the dynamics see Figure 6.1. Information is contained in the state description that encodes the differences among cell occupants in a study.

	A				B	
			A			
	B					B
		A				
B				A		

Figure 6.3. A cellular automata grid with occupied cells containing ingredients A and B.

The cell shape

The choice of the cell shape is based on the objective of the study. In the case of studies of water and solution phenomena, the square cell is appropriate since the water molecule is quadravalent to hydrogen bonding to other water molecules or solutes. A water molecule donates two hydrogens and two lone pair electrons in forming the tetrahedral structure that characterizes the liquid state. The four faces of a square cell thus correspond to the bonding opportunities of a water molecule.

The grid boundary cells

The moving cell may encounter an edge or boundary during its movements. The boundary cell may be treated as any other occupied cell, following rules that permit joining or breaking. A common practice is to assume that the grid is simulating a small segment of a very large dynamic system. In this model, the boundaries should not come into play in the results. The grid is then considered to be the surface of a torus shown in Figure 6.4.

In this case the planer projection of this surface would reveal the movement of a cell off the edge and reappearing at the opposite edge onto the grid as shown in Figure 6.5.

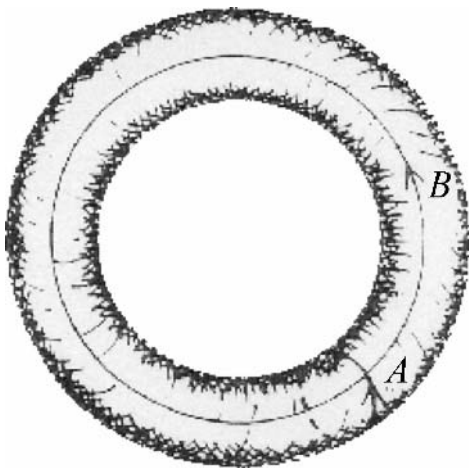


Figure 6.4. A two-dimensional grid mapped onto the surface of a torus.

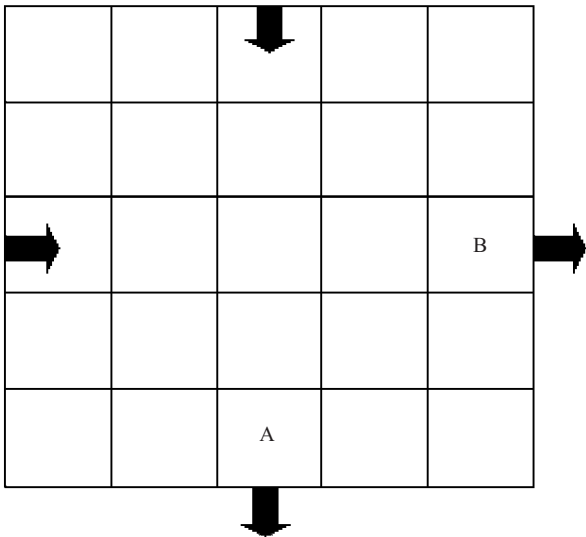


Figure 6.5. Paths of movement of ingredients on a torus, projected on a two-dimensional grid.

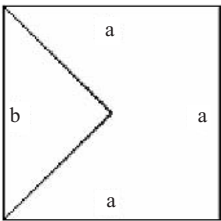


Figure 6.6. A variegated cell with two different sets of face states.

In some cases it is necessary to establish a vertical relationship among occupants. This establishes a gravity effect. For these studies the grid is chosen to be the surface of a cylinder with a boundary condition at the top and bottom which is an impenetrable boundary.

Variegated cell types

Until recently, all of the cellular automata models assumed that each edge of a cell had the same state and movement rules. Recent work in our laboratory has employed a variegated cell where each edge may have its own state and set of movement rules, Figure 6.6.

The cell shown in Figure 6.7 have three faces, *a*, with the same state and movement parameters while the other face, *b*, has a different state and movement parameters.

The three cells shown here have two faces, *a*, with identical states and movement parameters while the other two faces are different from *a* and from each other. Note that the faces, *a*, can be adjacent on the cell or they may be opposite.

Finally Figure 6.8 shows six arrangements of cell faces wherein all of the faces have different states and parameters. Note that mirror images or

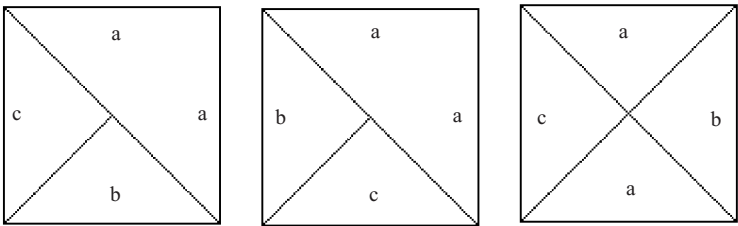


Figure 6.7. Variegated cells with three different face states.

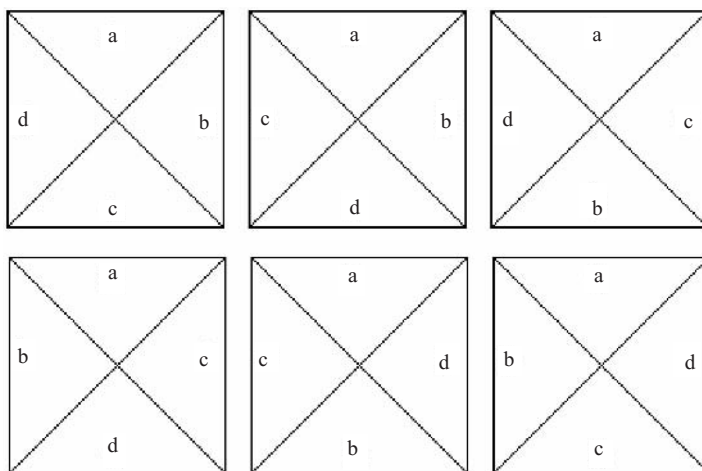


Figure 6.8. Variegated cells with four different face states.

chirality are present among these cells. Find the viral pairs. These variegated cells can be used for studies in which there are attempts to model different features within the same molecule.

3.3. Cell movement

The dynamic character of cellular automata is developed by the simulation of movement of the cells. This may be a simultaneous process or each cell, in turn, may execute a movement. Each cell computes its movement based on rules derived from the states of other nearby cells. These nearby cells constitute a neighborhood. The rules may be deterministic or they may be stochastic, the latter process driven by probabilities of certain events occurring.

Neighborhoods

Cell movement is governed by rules called transition functions. The rules involve the immediate environment of the cell called the neighborhood. The most common neighborhood used in two-dimensional cellular automata is called the von Neumann neighborhood, Figure 6.9, after the pioneer of the method.

Another common neighborhood is the Moore neighborhood, Figure 6.10, where cell, A, is completely surrounded by cells, B.

		B		
	B	A	B	
		B		

Figure 6.9. The von Neumann neighborhood. One cell, A, is in the center of four cells, B, adjoining the four faces of A.

	B	B	B	
	B	A	B	
	B	B	B	

Figure 6.10. The Moore neighborhood.

		C		
		B		
C	B	A	B	C
		B		
		C		

Figure 6.11. The extended von Neumann neighborhood.

Other neighborhoods include the extended von Neumann neighborhood shown in Figure 6.11, where the C cells beyond, B, are identified and allowed to participate in movements of the occupant of the A cell.

Synchronous/Asynchronous movement

When we speak of movement of a cell or the movement of cell occupants, we are speaking of the simulation of a movement from one cell to another. Thus a molecule or some object is postulated to move across space, appearing in a new location at time $t+1$. In the cellular automata models the actual situation is the exchange of state between two adjacent cells. If we are modeling the movement of a molecule from place A to the adjacent place B then we must exchange the states of cells A and B. Initially, at time t , cell A has a state corresponding to an occupant molecule, while adjacent cell B is devoid of a molecule, i.e., it is empty. At time $t+1$, the states of the cells A and B have exchanged. Cell A is empty and cell B

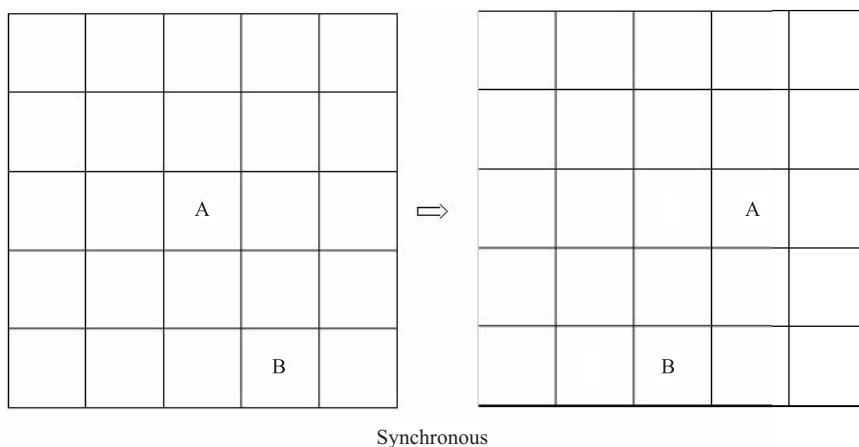


Figure 6.12. Synchronous movement of all ingredients in the grid.

has the state of the occupant molecule. This exchange gives the illusion, and the practical consequences of a movement of an ingredient from cell A to cell B. We speak of the movement of cells or of the movement of cell occupants; either way we are describing the process of simulating a movement as stated above. This effect is shown in Figure 6.12 for synchronous movement. Asynchronous movement is shown in Figure 6.13.

The movement of all cell occupants in the grid may occur simultaneously (synchronous) or it may occur sequentially (asynchronous). When all cells in the grid have computed their state and have executed their movement (or not) it is one iteration, a unit of time in the cellular automata model. In asynchronous movement each cell is identified in the program and is selected randomly for the choice of movement or not. The question of which type of movement to use depends upon the system being modeled and the information sought from the model but it should reflect reality. If the system being studied is a slow process then synchronous motion may be best represent the process. In contrast, if the system is very fast, like proton hopping among water molecules where the cellular automata is using a few thousand cells, then an asynchronous model is desirable. A synchronous execution of the movement rules leads to possible competition for a cell from more than one occupant. A resolution scheme must thus be in place

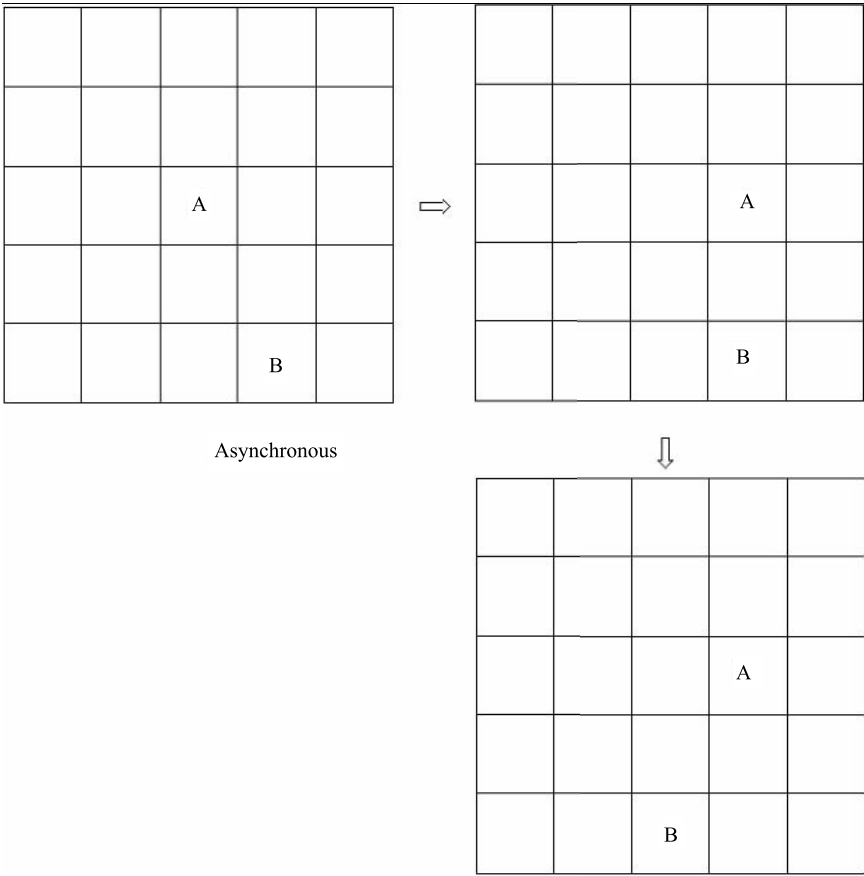


Figure 6.13. Asynchronous movement of two ingredients in the grid.

to resolve the competition; otherwise this may interfere with the validity of the model.

Deterministic/Probabilistic movement rules

The rules governing cell movement may be deterministic or probabilistic. Deterministic cellular automata use a fixed set of rules, the values of which are constant and uniformly applied to the cells of the same type. In probabilistic cellular automata, the movement of *i* is based on a probability-chosen rule where a certain probability to move or not to move is established for each type of *i* cell at its turn. Its state, (empty or occupied) is

determined, then its state as an attribute as an occupant is determined. The probability of movement is next determined by a random number selection between two predefined limits. As an example the random choice limits are 0 to 1000. A choice of numbers between 0 and 200 are designated as a “move” rule while a choice in the remaining number set, 201 to 1000, is a “no-move” rule; the case representing a probabilistic rule of 20% movement. Each cell then chooses a random number and behaves in turn according to the rule corresponding to that numerical value.

3.4. Movement (transition) rules

The movement of cells is based upon rules governing the events inherent in cellular automata dynamics. These are rules that describe the probabilities of two adjacent cells separating, two cells joining at a face, two cells displacing each other in a gravity simulation or a cell with different designated edges rotating in the grid. These events are the essence of the cellular automata dynamics and produce configurations that may possibly mirror physical events.

The free movement probability

The first rule is the movement probability, P_m . This rule involves the probability that an occupant in a cell, A, will move to an unoccupied adjacent cell. An example is cell A that may move (in its turn) to any unoccupied cell. As a matter of course this movement probability, P_m , is usually set at 1.0, which means that this movement always happens (a rule).

Joining parameter

A joining trajectory parameter, $J(AB)$, describes the movement of a molecule at, A, to join with a molecule at, B, or at C when an intermediate cell is vacant, shown in Figure 6.14.

This rule is computed after the rule to move or not to move is computed as described above. J is a non-negative real number. When $J = 1$, the molecule A has the same probability of movement toward or away from C as for the case when the C cell is empty. When $J > 1$, molecule A has a greater probability of movement toward an occupied cell B than when cell B is empty.

When $0 < J < 1$, the molecule A in Figure 6.15 has a lower probability of such movement. When $J = 0$, molecule A can not make any movement.

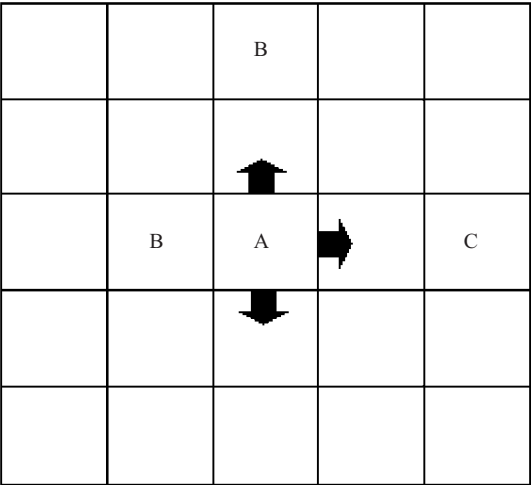


Figure 6.14. Results of a trajectory parameter operating on cell A ingredient.

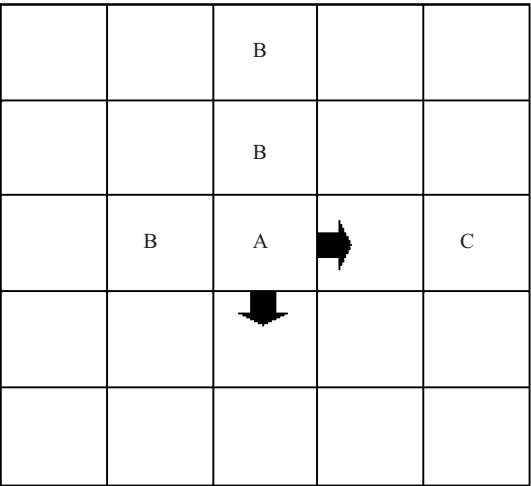


Figure 6.15. An arrangement on a grid where a movement away from grid C is favored when $0 < J < 1$.

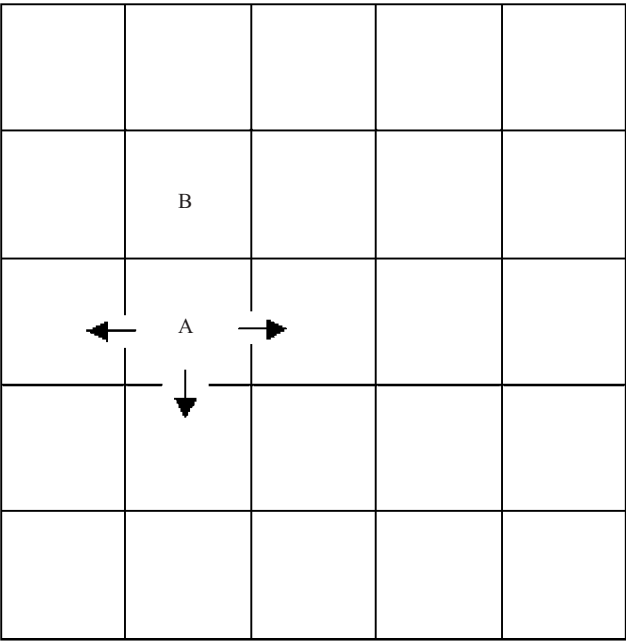


Figure 6.16. The breaking of A away from B governed by the P_B rule assigned.

Breaking rules

Just as two cells can join together, so their joined state can be broken. This is a movement governed by the trajectory rule called the breaking probability, P_B . To comprehend this it is convenient to refer to the occupant of a cell as a molecule. The $P_B(AB)$ rule is the probability for a molecule, A, bonded to molecule, B, to break away from, B, shown in Figure 6.16.

The value for $P_B(AB)$ lies between 0 and 1. If molecule A is bonded to two molecules, B and C, as shown in Figure 6.17, the simultaneous probability of a breaking away event from both B and C is $P_B(AB)*P_B(AC)$. If molecule A is bound to three other molecules (B, C, and D), shown in Figure 6.18, the simultaneous breaking probability of molecule A is $P_B(AB)*P_B(AC)*(P_B(AD))$. Of course if molecule, A, is surrounded by four molecules, Figure 6.19, it cannot move.

Relative gravity

The simulation of a gravity effect may be introduced into the cellular automata paradigm to model separating phenomena like the de-mixing of

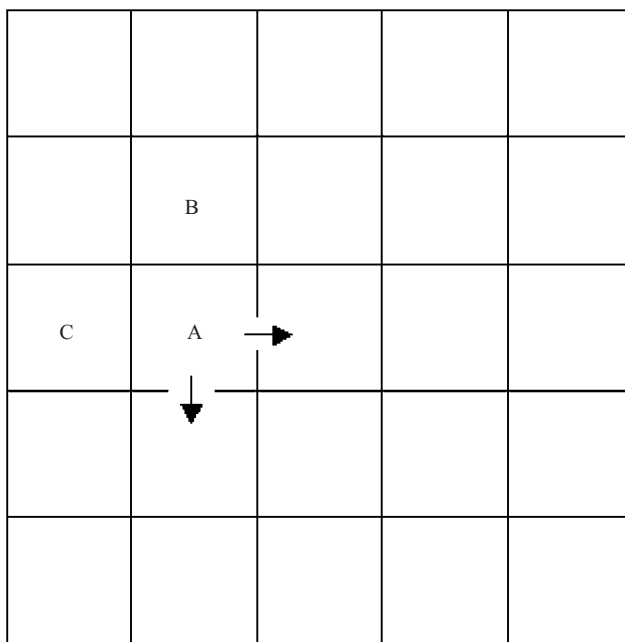


Figure 6.17. A grid arrangement where A is bonded to two other ingredients.

immiscible liquids or the flow of solutions in a chromatographic separation. To accomplish this, a boundary condition is imposed at the upper and lower edges of the grid to simulate a vertical versus a horizontal relationship. The differential effect of gravity is simulated by introducing two new rules governing the preferences of two cells of different composition to exchange positions when they are in a vertically joined state. When molecule A is on top of a second molecule B, then two new rules are actuated. The first rule $G_D(AB)$ is the probability that molecule, A will exchange places with molecule B, assuming a position below B. The other gravity term is $G_D(BA)$ which expresses the probability that molecule B will occupy a position beneath molecule A. These rules are illustrated in Figure 6.20.

In the absence of any strong evidence to support the model that the two gravity rules are complementary to each other in general cases, the treatment described above reflects the situations in which the gravity effects of A and B are two separate random events. Based on an assumption of complementarity, the equality $G_D(AB) = 1 - G_D(BA)$ may be employed in

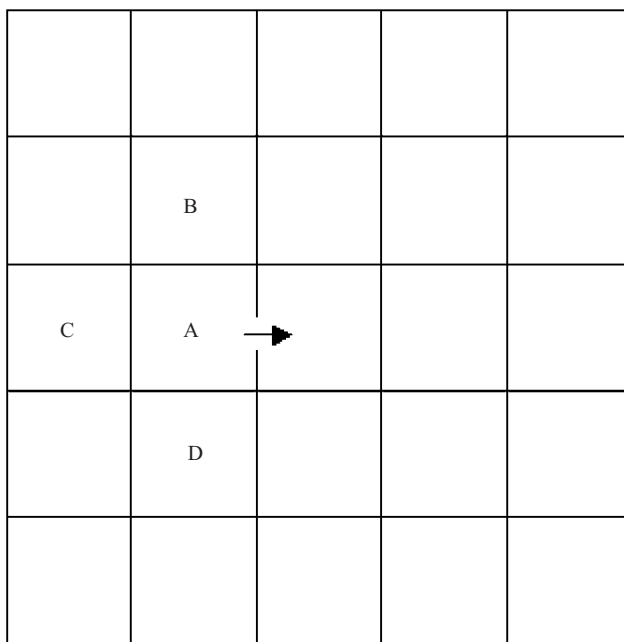


Figure 6.18. The grid arrangement where ingredient A is bonded to three other ingredients.

the gravity simulation. Once again the rules are probabilities of an event occurring. The choice of actuating this event made by each cell, in turn, is based on a random number selection from a set of numbers used for that particular event.

Absolute gravity

The absolute gravity measure of a molecule, A, denoted by $\text{absG}(A)$, is a non-negative number. It is the adjustment needed in the computation to determine its movement if, A, is to have a bias to move down (or up).

Cell rotation

In cases where a variegated cell is used it is necessary to insure that there exists a uniform representation of all possible rotational states of that cell in the grid. To accomplish this, the variegated cells are rotated randomly,

	B			
C	A	E		
	D			

Figure 6.19. Arrangement of ingredient A in which all four faces are bound to other ingredients.

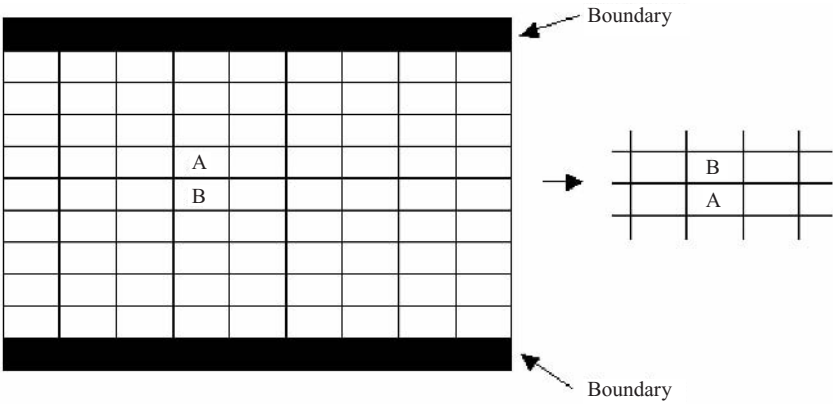


Figure 6.20. Sequence of movements reflecting the relative gravity rule.

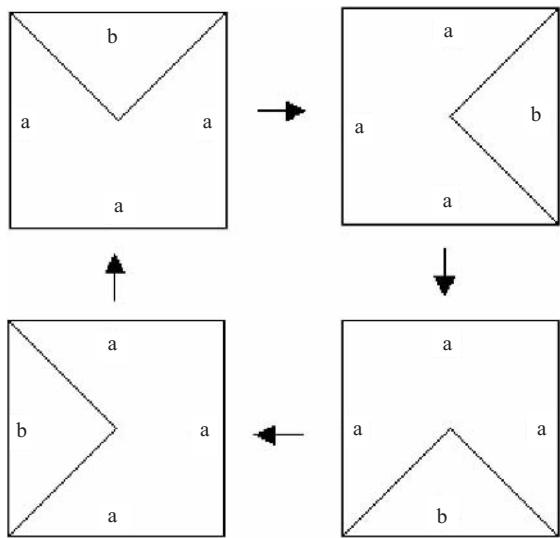


Figure 6.21. Cell rotations in four iterations.

(90°), every iteration after beginning the run. This process is illustrated for four iterations in Figure 6.21.

3.5. Collection of data

A cellular automata simulation of a dynamic system provides two classes of information. The first, a visual display may be very informative of the character of a system as it develops from initial conditions. This can be a dynamic portrayal of a process that opens the door to greater understanding of systems. The second source of information is in the counts of cells in different configurations such as those in isolation or those that are joined to another cell. This is called the configuration of the system and is a rich source of information from which understanding of a process and the prediction of unforeseen events may be derived.

Number of runs

It is customary to collect data from several runs, averaging the counts over those runs. The number of iterations performed depends upon the system under study. The data collection may be over several iterations

following the achievement of a stable or equilibrium condition. This stability is reckoned as a series of values that exhibit a relatively constant average value over a number of iterations. In other words there is no trend observed toward a higher or lower average value.

Types of data collected

From typical simulations used in the study of aqueous systems, several attributes are customarily recorded and used in comparative studies with properties. These attributes used singly or in sets are useful for analyses of different phenomena. Examples of the use and significance of these attributes will be described in later examples. The designations are:

f_0 - fraction of cells not bound to other cells

f_1 - fraction of cells bound to only one other cell

f_2 - fraction of cells bound to two other cells

f_3 - fraction of cells bound to three other cells

f_4 - fraction of cells bound to four other cells

In addition, the average distance of cell movement, the average cluster size and other attributes may be recorded.

4. Examples of Cellular Automata Models

4.1. Introduction

Over the past decade we have focused our attention on the use of cellular automata dynamics to model some of the systems of interest to the chemist and biologist. The early work in our laboratory has been directed toward the study of water and solution phenomena. This has resulted in a number of studies modeling water at different temperatures leading to a structural profile of degrees of bonding related to temperature. Such a profile is a structural surrogate for temperature in the correlations with properties of water. Properties normally related to temperature may now be related to structure. Studies include cellular automata models of water as a solvent [69], dissolution of a solute [70], solution phenomena [71], the hydrophobic effect [72], oil and water de-mixing [73], solute partitioning between two immiscible solvents [74], micelle formation [75], diffusion in water [76], membrane permeability [77], acid dissociation [78], and dynamic

percolation [79]. These studies have been summarized in reviews [80-82]. We will discuss some more recent cellular automata models carried out at the Center.

4.2. Water structure

Evidence shows that bulk water contains a significant amount of free space referred to as cavities or voids. It is obvious that water could not permit the diffusion of solutes through it if there was no space between water molecules. In ice, this is not the case since water molecules are bound to approximately four other water molecules. The choice of how many water molecules should be represented on a CA grid of a certain size was explored by Kier and Cheng [80]. Two approaches were taken. On the basis of estimates of the volume of a water molecule and the number of water molecules in a mole, an estimate of about 69% occupancy of a grid was deduced.

The second approach was to conduct CA runs with varying water concentrations. The attributes of the CA configuration were interpreted and compared with experimental values. After a sufficient number of runs, the average number of cells joining each cell was recorded. This attribute was judged to be a model of the average number of hydrogen bonds per molecule of water. A good correspondence of this value was found for a water concentration of about 69% of the grid cells. Another attribute from these experiments, the number of free, unbound water molecules was recorded. This small percentage of the total number of waters was compared with the number of free waters from experiment. The best correspondence was found for a CA system containing about 69% water molecules in the grid. From this information, a system modeling water was adopted using 69% water in the grid.

Water movement rules

Three rules must be chosen to impart a “water character” to the occupied cells that we designate. The first of these is the movement probability P_m . There is no practical reason to believe that anything other than $P_m = 1.0$ for water has any real significance. This choice characterizes water as a freely moving molecule whenever it is possible. The other two rules governing the joining and breaking of water molecules are critical to their behavior and to the emergent attributes of the CA dynamics. Recall that the joining

rule, J , encodes the probabilities of water molecules to join others to form a bond (a hydrogen bond in the case of water). The breaking probability, P_B , describes the tendency of bound waters to break apart. The selection of these rules is essential if the model is to have any validity.

The linkage between rules J and P_B can be made relative to a range of values of one of them. As described earlier, the P_B value ranges from zero to one, therefore the J value may be chosen as a function of P_B .

$$\text{Log } J = -1.5 P_B + 0.6 \quad (4.1)$$

The wisdom of this choice can be tested by comparing the attributes from the dynamics with physical properties.

The attributes recorded

A CA run leads to a configuration that is constant in an average sense. Several attributes of this configuration may be recorded and used for further

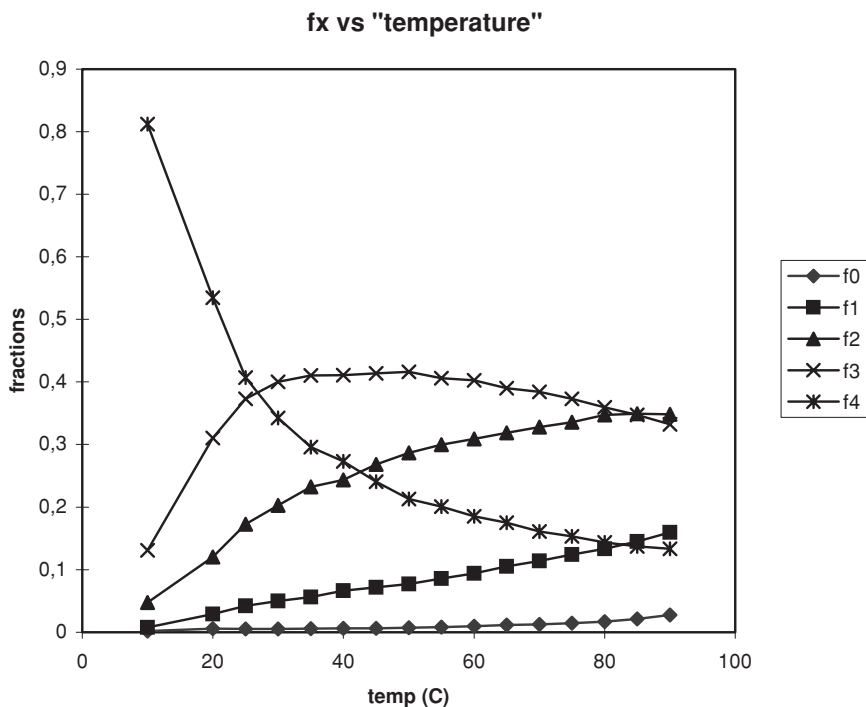


Figure 6.22. Fractions of cell binding states as a function of modeled temperature.

Table 6.1. Water properties related to cellular automata attributes*

Property	Equation	r ² -correlation
Vapor Pressure	$\text{Log } P_v \text{ (mm Hg)} = 13.77(f_0 + f_1) + 0.795$	0.987
Dielectric Constant	$\epsilon = -224 f_1 + 86.9$	0.989
Viscosity	$\eta \text{ (centipoise)} = 3.165 f_4 - 0.187$	0.989
Ionization	$-\text{Log } K_w = -20.94 f_H + 16.43$	0.999
Surface Tension	$\gamma \text{ (dynes/cm)} = 16.07 N_{HB} + 22.35$	0.970
Compressibility	$\kappa \text{ (}\times 10^6/\text{Bar)} = -53.82 f_3 + 66.66$	0.953

* f_H is the average number of free hydrogens per water molecule. N_{HB} is the average number of bonded neighbors per water molecule.

study. The configuration may be analyzed for the numbers of molecules with no neighbors, one neighbor, and so on up to four neighbors. The fraction of each state are represented as shown in the f_x vs “Temperature” plot above. The distribution of these fractional values becomes a profile of the state of a molecule. It is observed that these states change with different J and P_B rules and that these states have some correspondence to the temperature of the system. The f_x attributes computed for different values of P_B and the corresponding J rules are plotted in Figure 6.22.

If we assign a “temperature” of the water to a P_B value according to the relationship:

$$t(^{\circ}C) = 100 P_B \tag{4.2}$$

then sets of f_x values can be related to temperatures of water. From this relationship, selected physical properties at different temperatures may be related to f_x values at those temperatures. An analysis of several properties demonstrate the relationship with selected f_x values and the general validity of our CA model of water [78,79]. Some of these analyses are shown in Table 6.1.

The conclusion that selected values of the movement rules produce a meaningful profile of f_x values, makes it possible to proceed with some confidence that this CA model of water has validity.

4.3. Cellular automata models of molecular bond interactions

One emergent property arising from an ensemble of agents in a complex system is the extent of molecular interactions, which is the pattern of behavior as free moving agents join and break with their neighbors. In

the condensed phase, molecules like water move about, bind with each other in response to input of energy usually in the form of heat. A vigorous pattern of repulsive interactions leads to the transition of the liquid to the gaseous state. The extent of these interactions may be modeled by recording the encounters of molecules with a thermometer bulb. A displacement of mercury in the thermometer is taken to be a model of a certain number of interactions among the liquid molecules. At the phase transition the recorded temperature is the boiling point. This emergent property is a consequence of the structure of the molecules exhibiting this behavior. What is it about the structure difference between two different molecules that gives rise to two different boiling points? It is the topological structure of the molecule, permitting more or fewer binding interactions among molecules.

The greater the number of binding interactions, the greater the tendency of the molecules to remain in contact with a neighbor. This translates into a lower propensity of the molecules to be liberated from the bulk liquid and to remain as an ingredient in the liquid. The fewer the intermolecular bond interactions, the lower the recorded temperature. If we could reckon the relative extent of these binding states, we may have a structural model of the water at the temperature of the boiling point.

A novel approach to modeling intermolecular interactions was proposed by Kier [81] in which encounters of molecules were dissected into the encounters of individual bonds, an approach called *dissecta membra*. Each bond type was simulated by an occupied cell on a cellular automata grid. Each occupied cell has a particular state value, derived from the descriptors, C_{ij} in Table 6.2.

The rules for joining and breaking of each type of CA cell with another CA cell were derived from the bond encounter possibilities using the product of the bond types in the encounter, $(C_{ij})(C_{kl})$, between two molecules. To scale these possibilities to the trajectory J rules for our cellular automata model we have adopted the relationship:

$$J \text{ for } (C_{ij})(C_{kl}) = 4(C_{ij})(C_{kl}) \quad (4.3)$$

The breaking probabilities, $P_B(C_{ij})(C_{kl})$, were derived from the relationship used in studies of water:

$$\log J(C_{ij})(C_{kl}) = -1.5 P_B(C_{ij})(C_{kl}) + 0.6 \quad (4.4)$$

Each carbon-carbon bond in an alkane was represented by a particular state of a cell. One hundred molecules were modeled in a grid of 3025 cells. For example, the *dissecta membra* model of 3-methyl pentane uses

Table 6.2. Alkane bond types and connectivity descriptors, C_{ij}

Bond type, $\delta_i - \delta_j^{(a)}$	$C_{ij} = (\delta_i \delta_j)^{-0.5}$
1,1	1.000
1,2	0.707
1,3	0.577
1,4	0.500
2,2	0.500
2,3	0.408
2,4	0.354
3,3	0.333
3,4	0.289
4,4	0.250

^(a) The symbols δ_i and δ_j are the counts of the number of carbon atoms bonded to atoms i and j . Atoms i and j are bonded to each other.

200 cells with a state corresponding to bond type (1,2), 100 cells with a state corresponding to bond type (1,3), and 200 cells with a state corresponding to bond type (2,3). Each cell moved randomly during one iteration, joining another cell, breaking from another cell or moving freely in unoccupied grid space. The dynamics were run for 990 iterations and then during the next 10 iterations the count of the number of joined cells was recorded. This process was repeated for 25 runs and the count of joined cells was averaged from this data. The count of joined cells was called the beta, β , value. Thirty eight alkanes including all of the pentanes, hexanes, heptanes and octanes and three cycloalkanes were modeled and the beta values recorded

Results

The interpretation of the β value is a relationship with a physical property that is highly dependent upon intermolecular binding interactions. One such property is the boiling point B. P.). The β values, expressed in a quadratic equation, relates to the boiling point with the statistics shown:

$$B.P. (^{\circ}C) = 0.584 \beta - 0.0004 \beta^2 - 71.517$$
$$r^2 = 0.991, s = 2.996, n = 38, F = 2027$$

(4.5)

Discussion

The count of cell encounters averaged over time (iterations), encoded by the beta value is very closely related to the boiling point of the

38 alkanes. The standard deviation of only 3.00 degrees is better than any one-variable quadratic analysis we have found reported. The cellular automata dynamics improves the modeling of bond encounters compared to the static count of all possible bimolecular encounters. The quality of the relationships revealed here supports the cellular automata dynamic model using the *disject membra* simulation of the important topological features of these molecules.

We propose that treating the bonds of a molecule as *disjecta membra* is a model with a limited objective and a limited relationship to reality. It is an example of the analysis of a complex system using reduction to isolate relevant parts followed by synthesis using cellular automata dynamics to create a model that reveals some information about emergent properties and the role that the ingredients contribute to the whole. In this case, we considered just the bonds in alkanes, endowed with numerical values reflecting their accessibility to other bonds in other molecules. Our dynamic model in a limited way, simulates the conditions of a molecule in its milieu.

4.4. Diffusion in water

The diffusion of solutes through water is a means of transport of both vital and noxious compounds within the body. We see this phenomenon everywhere, across a synapse, through a nephron, within cells, along blood vessels; where water goes, so go the solutes via diffusion. Water is uniquely structured to facilitate this passage by virtue of its reactivity in bulk. A water molecule will form up to four hydrogen bonds with neighbors. This is a dynamic pattern, with hydrogen bonds forming and breaking at a rate of 10^{-14} seconds. Water molecules constitute about 2/3 of the space they occupy in bulk, thus large volumes of space are available for solute passage as the architecture of the water changes. It is clear that these spaces, voids or chreodes are the passages through which all diffusion occurs through water.

The term, chreodes was used by Kier [82] in a recent article describing a theory of the facilitated and preferred passage of ligands across the landscape of a protein to a receptor or enzyme active site. The chreode is a temporarily connected series of evanescent cavities in the water enshrouding the field of amino acid side chains on the surface of a protein. The varying hydrophobic states of these side chains produces an influence on the nearby bulk water ranging from the hydrophobic structuring of water to electrostriction. This varying pattern creates evanescent, favored pathways,

chreodes, through the water whereby a molecule experiences a facilitated diffusion from anywhere on the protein surface, to the receptor.

The influences on the molecule that govern its diffusion characteristics are its size, its hydrophatic state, and the temperature (structure) of the water. Some experimental evidence and modeling has demonstrated these influences, leading to the conclusion that more hydrophobic molecules diffuse faster than hydrophilic molecules of the same size [73]. The studies of these influences are scarce because of the difficulties in conducting diffusion measurements with varying parameters among the ingredients. The modeling results mirror reality as far as the hydrophatic state influence on the rate of diffusion. A major advantage of these kinds of models is that it is possible to manipulate one variable while holding others constant, thus creating a profile of a system and its behavior under several sets of conditions.

In this study we examined by modeling, the influence of water temperature and solute hydrophatic state on the diffusion of the solute through water. In addition, we modeled the water and the cavities within it to attempt to explain the observed behavior. The modeling is accomplished using asynchronous, probabilistic cellular automata, as we have described above.

Temperature and hydrophatic state influences on diffusion rate

In order to establish the relationship between the extent of diffusion and the water temperature, we first recorded the diffusion every 100 iterations up to 1000 iterations for several modeled temperatures of water, labeled W. The hydrophatic states of the solute molecules, labeled S, were held constant at an intermediate value. This was accomplished by using the parameters for solute-water interaction as $P_B(W,S) = 0.5$ and $J(W,S) = 0.7$. This modeled a solute with a mid-level hydrophatic state. The solute-solute interaction parameters were held constant at $P_B(S,S) = 0.5$ and $J(S,S) = 0.7$. The diffusion was shown to be linear with time (iterations), characteristic of a random walk where the distance traveled is known to be linear with a function of time. These results produced a confidence in the model and led us to the use of a common iteration time, 1000 iterations, to compare various hydrophatic states with their influence on diffusion.

The extent of diffusion at 1000 iterations was then recorded for various combinations of water temperature and solute hydrophatic state. These results are shown in Table 6.3.

Table 6.3. Extent of diffusion at 1000 iterations

Hydropathic state	Temperature						
$P_B(\text{WS})$	0.10	0.20	0.30	0.40	0.50	0.60	0.70
0.10	0.50	0.54	0.29	0.20	0.14	0.19	0.21
0.20	1.24	1.75	1.25	1.20	1.15	1.08	1.20
0.30	1.74	3.40	3.74	4.02	3.44	3.13	3.25
0.40	2.05	3.60	5.24	4.78	5.61	5.52	4.34
0.50	2.68	3.90	4.18	6.22	6.96	7.11	7.02
0.60	2.36	4.14	6.42	7.85	7.56	8.95	8.45
0.70	2.11	3.92	7.08	8.02	9.36	9.49	9.27

The diffusion was measured as the average count of cells traversed by the solute from the center of the grid after 1000 iterations. This average was computed from 40 runs. The distance was counted as cell faces, not diagonal distances since the solutes only move in rectangular patterns.

Results

From this table we see that diffusion is modest at all temperatures for relatively hydrophilic solutes. As the hydrophobic character increases, the diffusion rate increases for all temperatures. The response to these parameters is, however, not linear. As can be seen, the diffusion of mid-level hydrophobic solutes goes through a maximum at mid-temperatures. When the solute hydropathic state is non-polar the diffusion increases increased temperature. With intermediate hydropathic states, modeled with $P_B(\text{WS}) = 0.3, 0.4,$ and 0.5 , the diffusion rate is maximal at intermediate temperatures modeled with $P_B(\text{W}) = 0.4, 0.5,$ and 0.6 , falling to lower rates at higher temperatures.

We surmise that this non-linear behavior must be due to at least two intersecting changes in the complex system formed by the solute and water at different temperatures and hydropathic states. To explore these possibilities we have modeled the water architecture at different states relevant to conditions described above.

The architecture of water

Another effect emerging from temperature changes of water is the distribution of the cavities among the molecules. A study of the average cluster size and distribution of the clusters may shed some light on its behavior

Table 6.4. Average Cavity Structure

$P_B(W)$ (Water temperature)	Average cavity cluster size (cells)
0.20	7.2
0.30	4.2
0.40	3.0
0.50	2.4
0.60	2.1
0.70	1.8
0.80	1.6

in the presence of solutes of differing hydropathic states. We have run cellular automata simulations of water at various temperatures, recording the average size of the cavities. These simulations treated the cavity spaces as discrete entities and so we could make the measurements of their average size and distribution. Table 6.4 reveals the architecture of water cavities at various temperatures.

Discussion

The rate of diffusion declines at a certain temperature because the increase in the number of hydrogens available for hydrogen bonding increases. This produces a greater extent of solute hydration, which produces a larger effective size hence, resistance to diffusion through the cavities. As the temperature rises from cold to about mid-temperature, the average size of the water cavities decreases and openings between cavities occur. This permits diffusion of a solute to occur more readily between groups of cavities. As the temperature rises above the mid level, the cavities become smaller in size but more connected. As a result, there is more cavity surface area exposed to solutes. If the solute is hydrophilic, it will form more hydrogen bonds with this increased polar surface of the cavity. These hydrogen bonds produce effectively larger, hydrated solute molecules, reducing the diffusion rate. These two intersecting effects produced a non-linear rate of diffusion for solutes in the mid-range hydropathic state.

4.5. Chreode theory of diffusion in water

A major focus of attention in drug-receptor studies is the structural influence on the encounter of these two molecules. Variation in the structure of the drug is often related, through models, to the binding affinity with

a receptor or the catalytic rate with an enzyme. This leads to progress in the design of new, active biomolecules and some understanding of the processes underway. Less attention has been paid to the journey of the drug to the receptor.

The role of diffusion

An early view described the approach of a drug through bulk water to a receptor. This three-dimensional random walk is postulated by Welch [83] and others to be too slow to permit coordination of the heterogeneous metabolic processes in living systems. The current view invokes some period of residence of the drug in the vicinity of the protein surface, on the way to an encounter with a receptor. The nature of this two-dimensional passage is not generally agreed upon and is the subject of several models. The central idea is that the approach of a drug to a receptor is a process assumed to be diffusion, acting as a limiting condition to the rate of the reaction [84-86].

It is well-known that water is an essential ingredient in the reactions of biological phenomena, indeed the complex system of drug-receptor-water is described by Kier and Testa [87] as a triad that is at the core of biological transformations. Any model of two-dimensional diffusion of a drug across a membrane or protein surface must include the participation of water. Water is an evanescent substance, constantly changing its architecture by making and breaking hydrogen bonds. Diffusion through water is a passage through the voids between clusters of hydrogen-bonded water molecules. The presence of solute molecules has a varying influence on water molecules in their vicinity depending on the interactions possible between the different species. Polar solutes form hydrates, collections of water molecules in intimate contact with the solute molecule. In contrast, non-polar solute molecules are not effectively bound to water molecules, allowing the local organization of water to occur, driven by the preference of water to bind to water rather than water binding to solute. This phenomenon is called the hydrophobic effect.

Evidence may be interpreted as describing a facilitation of reaction rates and receptor activation due to the rapid diffusion of molecules to target sites in essentially a two-dimensional domain. This diffusion most likely occurs across the surface of a protein, involving structural features not part of the receptor. The process of guidance of the drug molecules to the receptor would be expected to result in several effects. These include some period

of retention on the protein surface, a minimum extent of binding delaying the passage, and some influence facilitating the movement of the drug to the receptor. These considerations and the demonstrated role of water led us to consider a model involving the immediate layers of water enshrouding the protein.

Drug molecule diffusion and the hydrophobic effect

We have proposed that drug molecules encounter the surface of a protein molecule and are captured within the first few layers of water on the surface. They are then guided to the receptor over a series of cavity paths in the water created by the hydrophobic effect responding to the hydropathic state of each amino acid side chain [84]. These paths are preferences reminiscent of the chreodes envisioned by Waddington [88] in his description of an epigenetic landscape. He coined the word chreode from the Greek words for “necessary” and “route” or path. He defined it as a representation of a temporal succession of states of a system. That system is characterized by a property that a dynamic system will tend to respond to perturbations by returning to the chreode. There are two characteristics encoded in this concept. The first is the presence of a degree of progress as one proceeds from the initial to the final state of the system. In some periods of the traverse through the chreode, there is a great deal of progress; in other periods, less so. The second characteristic is the relative strength of the tendency of the ingredient to return to the trajectory created by a chreode. Because this definition is close to the phenomenon Kier adopted this term to characterize the system.

The hydrophobic effect arises from the greater attraction and binding of water to itself, rather than water binding to another molecule (a solute or a stationary molecular fragment). The relative hydropathic state of the other molecule influences the degree of aggregation of the water molecules in the vicinity. On the surface of a membrane or protein, Welch [89] referred to as the microviscosity. To reveal this effect, we have previously calculated the relative hydrophobicity of a solute in water in a cellular automata simulation [90]. The hydropathic state is encoded in the rule, $P_B(WS)$, where a high probability reflects a hydrophobic solute and a low value reflects a hydrophilic solute. This earlier study illustrated the ability of the hydropathic state of a solute to organize the water in its vicinity. From this, we infer that the amino acid side chains, with variable hydropathic states, may

organize the water and the cavities into some pattern which could function as our postulated chreode.

The hydrophobic state and ligand diffusion

Our previous studies have shown that the diffusion of a solute through a solution is influenced by the hydrophobic state of other solutes [91]. The diffusion of a solute was demonstrated to be faster if the other solute is hydrophobic. In our model of a chreode on a hydrodynamic landscape, it is necessary to consider the influence on diffusion of multiple stationary ingredients representing the side chains on a protein surface.

Creation of a model chreode

With the information gleaned from the gate studies and hydrophobic influences, we next explored the possibility of a guided or directed trajectory for the diffusant; in essence a model of a chreode on the hydrodynamic landscape [84]. For this study, we used a cellular automata grid located on the surface of a torus with a central region representing a target for a ligand, shown in the figure.

A random distribution of the five cell types representing the five groups of amino acid side chains based on hydrophobic state, Figure 6.23, were introduced onto a cellular automata grid. All of these cells, (A,B,C,D,E) were separated from each other by 3 cell spaces, not explicitly shown but present in the calculation. The system contains water in the same proportion as used in the previous studies. The water is allowed to move freely and to interact with the ligand and stationary cells representing amino acid side chains of various hydrophobic states. The diffusant was started at a position 38 cells from the target and endowed with a neutral hydrophobic state in this study. The count of iterations necessary for the ligand, S, to traverse the grid and touch the central target was averaged over 200 runs. These values and those in the following study had a standard error of the mean of 6%. The number of iterations necessary in this study to traverse the distance to the center was calculated to be 12,429.

A second model was created in which the same number of A,B,C,D, and E structures were randomly scattered, Figure 6.24. Each cell was separated from another by a 3-cell space. Eighteen of the 100 cells were organized to form a chreode pattern shown in italics. In this pattern the order of increasing hydrophobic character is assigned in the order A,B,C,D,E.

S	C	C	D	D	B	E	B	C	D
D	B	E	B	C	C	A	C	D	C
B	D	A	D	A	C	D	B	B	D
A	C	B	A	D	A	A	C	E	A
B	D	C	E	B	E	C	B	A	C
E	C	A	D	♥	A	B	E	B	B
C	A	C	A	C	E	A	C	D	C
C	A	B	D	A	E	C	B	A	D
A	D	A	B	B	A	B	A	C	B
C	A	C	A	D	B	A	E	A	C

Figure 6.23. A grid representing the protein landscape of a receptor, ♥. The side chains represented by the letters are randomly scattered over the grid. The water moves freely among them. The ligand, S, is allowed to diffuse among the side chains until it reaches the receptor.

S	C	C	D	<u>A</u>	B	E	B	C	D
D	B	E	B	<u>B</u>	C	A	C	D	C
B	D	A	D	<u>C</u>	C	D	B	B	D
A	C	B	A	<u>D</u>	A	A	C	E	A
B	D	C	E	<u>E</u>	E	C	B	A	C
<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	♥	<u>E</u>	<u>D</u>	<u>C</u>	<u>B</u>	<u>A</u>
C	A	C	A	<u>E</u>	E	A	C	D	C
C	A	B	D	<u>D</u>	E	C	B	A	D
A	D	A	B	<u>C</u>	A	B	A	C	B
C	A	C	A	<u>B</u>	B	A	E	A	C

Figure 6.24. A grid representing the protein landscape of a receptor, ♥. The side chains represented by the letters are randomly scattered over the grid. The water moves freely among them. A number of side chains are arranged in an order of increasing hydrophobicity around the receptor. The ligand, S, is allowed to diffuse among the side chains until it reaches the receptor.

The dynamics were run as before and the time for the S cell to traverse the spaces to the central target was averaged over 200 runs. In this study the average time to diffuse to the center was calculated to be 8,609 iterations. This grid models the possible diffusion of a ligand across a surface with specifically positioned stationary cells coordinating their hydrophobic states to facilitate diffusion toward the center of the grid. This models a chreode as we have defined it in this study. It was found that the rate of diffusion of a ligand is faster in the ordered stationary cell model as compared to a random distribution of these same amino acid side chain cells on the grid. The diffusion observed in the second study is a consequence of the existence of a pattern of cells and their hydrophobic states, a possibility existing on the surface of a protein.

Discussion

The diffusion of ligands across protein surfaces has been considered as a route for these molecules to reach the active sites. This model is offered to explain the rapid response of receptors and the fast rates of enzyme catalysis. Two general mechanisms have been proposed in the past to define the facilitation of the diffusion. In the first case, the forces of interaction between ligands and the side chains are suggested to be electrostatic. If this were true the attraction between ligand and certain side chains might be strong enough to ensnare the molecule in regions of the protein surface, retarding diffusion. These side chains would, in essence, function like an active site or like a binding site rather than a diffusion-promoting feature. In contrast, van der Waals forces have been proposed between ligands and side chains, serving to facilitate the diffusion to the active site. These forces require a very close approach of the ligand and side chain, presenting a deterrent to rapid diffusion because of steric entanglement.

A theory proposed in this study invokes the participation of the layers of water molecules immediately adjacent to the protein surface [84]. Each amino acid side chain intruding into the bulk water exercises an influence on the water architecture. This takes the form of a hydrophobic effect from hydrophobic side chains and side chain hydration with hydrophilic side chains. The cavities between clusters of hydrogen-bonded water molecules are in a dynamic state, joining with other cavities, breaking away, reforming, all in a manner resembling a dynamic peristaltic pump. These cavities form the proposed chreodes facilitating the passage of diffusants across the protein surface.

Several consequences of the theory may be derived from a consideration of known phenomena. The measurement of the affinity of a molecule may reflect a more complex system than just a drug-receptor encounter. It may be that the measurement is including some residence of the molecules in chreodes. Another observation that may have a connection with the chreodes is the phenomenon of lag, where some time must pass before an effect is observed in a pharmacological test system. The lag may reflect the time needed for the drug molecule to displace the transmitter from the chreodes leading to the active site. The sequel to that observation is persistence, where the measured effect continues for some time after the washout of the ligand from the test system. The velocity of enzyme reactions may, in part, be explained by the facilitated trajectory of the ligand through chreodes to the active site and the velocity of departure of the product through chreodes. The chreodes may exhibit some selectivity of their occupants and may possibly reject some molecules that are detrimental to fitness of the system. Finally, the mechanism of non-specific anesthetic agents may depend on their ability to interfere with the chreodes carrying the normal transmitter to the receptor.

A theory of volatile anesthetic action

A theory was proposed by Kier [92] for the clinical actions and behavior of the volatile anesthetic agents. Evidence supports the non-specific and ubiquitous effects of these drugs in which many ligand-receptor systems may be involved. The drugs interfere with the diffusion of ligands to receptors across the hydrodynamic landscape on the protein surface. The hydropathic states of the amino acid side chains surrounding an active site influence the water to form chreodes. These chreodes are altered by the presence of the volatile anesthetic drugs, leading to a reduction in the diffusion rate of numerous ligands to their active sites. This effect is sufficient to decrease the responses of numerous receptors, leading to responses characteristic of the anesthetic states.

4.6. Modeling biochemical networks

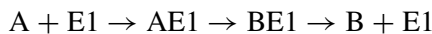
Dynamic evolutionary networks have recently been recognized as a universal approach to complex systems, ranging from quantum gravity to biological cells and organisms, ecosystems, social groups, and market economy. The network approach is a non-reductionist approach enabling

analysis of the systems as a whole, which makes it an ideal tool for systems biology. Network topology is generally used in characterizing networks, focusing on their connectivity, neighborhood and distance relationships. Network complexity has also been recently quantitatively characterized [93]. This abundance of cellular networks data, produced by microarrays, 2D-gel chromatography, mass-spectra, and other techniques, brings about another dimension of the network approach, allowing the tracing of the continuous changes of network species and their interactions. The large size of the metabolic, protein, and gene regulatory networks makes impractical many of the traditional methods for dynamic modeling. It is the purpose of this study to outline the potential of cellular automata as a basic method for the dynamic modeling of networks for biological and medical applications [94].

The MAPK cascade signaling pathway

We have begun the analysis of a protein network [94], selecting the mitogen-activated protein kinase (MAPK) cascade as an example of importance, studied recently by numerical solving of the reaction rate equations [95]. This is a signaling pathway, relaying signals from the plasma membrane to targets in the cytoplasm and nucleus. The cascade is shown schematically in Figure 6.25.

The first reaction of the MAPK cascade implies the detailed reaction mechanism shown here:



where A stands for MAPKKK, and B for MAPKKK*. A product of a reaction may be a molecule, cell, etc., but it may also be a catalyst or enzyme. In the MAPK cascade, this is the case with the activated MAPKKK and the MAPKK-PP, which are catalysts for the forward reactions in the second and third cascade level, respectively. The three substrates in the cascade have some prescribed initial concentration.

It is assumed that the enzymes have considerably smaller concentration than the substrates. The enzymes are modeled as being selective, each for a specific encounter (complex) and a specific reaction outcome. Finally, the model requires that each protein (but not necessarily each enzyme) is free to move about and to encounter other proteins and enzymes. The selectivity of the enzymes determines the conversion of one protein to another in a discrete way, thus the network display encodes these encounters and outcomes

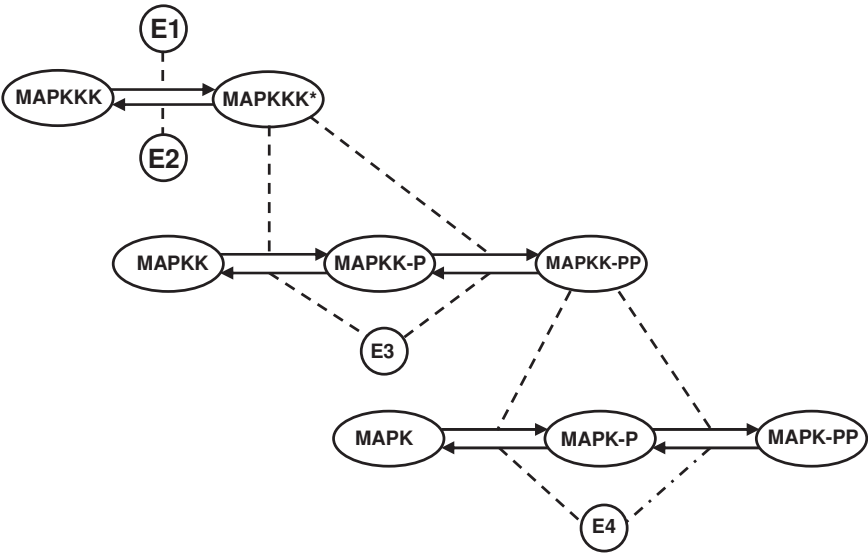


Figure 6.25. The MAPK signaling cascade. The substrates and products are represented by oval contours, reactions by arrows, and the catalysts action by dashed lines. E3 and E4 are MAPKK-protease and MAPK-protease, respectively. P stands for phosphate, PP for diphosphate.

The question asked in an analysis of a network such as Figure 6.25 is, what is the consequence of the change in a concentration of a protein or an enzyme in the system. The change in the substrates concentrations changes directs the reversible reactions toward the desirable outcome. The change in an enzyme concentration changes the rate with which the reaction reaches the equilibrium state. However, in non-equilibrium reactions, which frequently occur in biochemistry, different enzyme concentrations, produce a different steady-state, thus influencing the degree of conversion of substrates into products. One approach to these questions and to describe the patterns of behavior of the dynamic network is to model the system using cellular automata (CA). We will follow the general method used by Kier [96,97] in setting up a CA model of enzyme activity.

The CA modeling design

Each molecule involved in the MAPK pathway is represented by a number of cells in the CA grid. The numbers chosen reflect the relative concentration of that protein. Each of the cells representing all other

proteins moves about freely in the grid. They may encounter each other but this has no consequence. The only encounters that have a consequence are those between a specific protein (substrate) and a specific enzyme, as shown in the network. When such an encounter occurs, there is modeled a complex (enzyme-substrate). This complex has an assigned probability of converging to a new complex (enzyme-product). Following this there is a probability assigned for the separation of these two species.

Our studies of the MAPK cascade were performed using a CA grid of 100 by 100 cells. Each model was obtained as the average of 50 runs, each of which included 5000 iterations, a number sufficiently large to enable reproducing the steady-state of reaction. The grid used had no boundary conditions, thus movement past an edge puts the substance at the opposite face. A network to be studied is represented by groups of CA cells, each group representing one of the network species. The number of cells in each group reflects the relative concentrations of each network ingredient. We have systematically altered the initial concentrations of several proteins (MAPKKK, MAPKK, and MAPK) and the competencies of several enzymes (MAPKK- and MAPK-proteases, and the hypothetical enzymes E1 and E2 that affect the forward and reverse reactions of activation and deactivation of MAPKKK). The basic variable was the concentration of MAPKKK, which was varied within a 25-fold range from 20 to 500 cells. The concentrations of MAPKK and MAPK were kept constant (500 or 250 cells) in most of the models. The four enzymes, denoted by E1, E2, E3, and E4, were represented in the CA grid by 50 cells each. In one series of models, we kept the transition probabilities of three of the enzymes the same, ($P = 0.1$), and varied the probability of the fourth enzyme within the 0 to 1 range. In another series, *all* enzyme probabilities were kept constant, whereas the concentrations of substrates were varied. The last series varied both substrate concentrations and enzyme propensities. Recorded were the variations in the concentrations of the three substrates MAPKKK, MAPKK, and MAPK, and those of the products MAPKKK*, MAPKK-P, MAPKK-PP, MAPK-P, and MAPK-PP.

Modeling enzymes activity

Upgrading or downgrading enzymes activity is one of the typical ways the cell reacts to stress and interactions with pathogens. We studied systematically the variations of one of the four enzymes E1 to E4 at constant

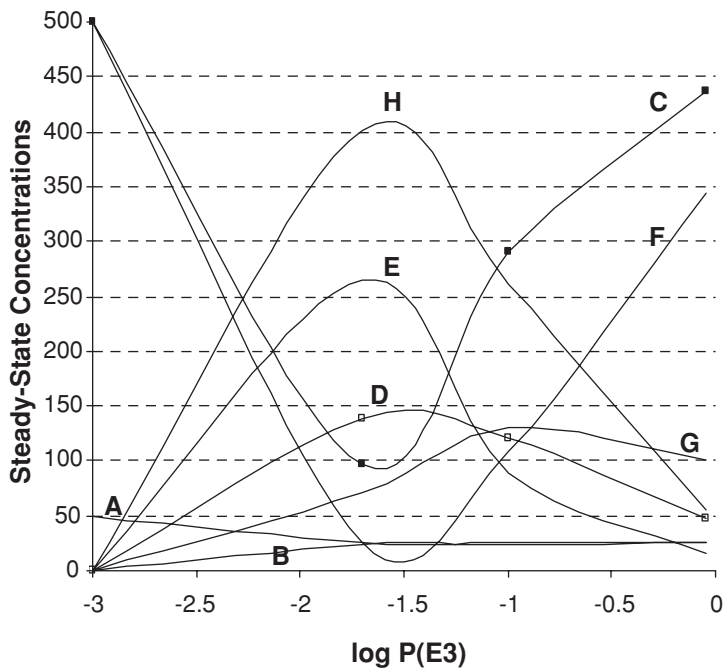
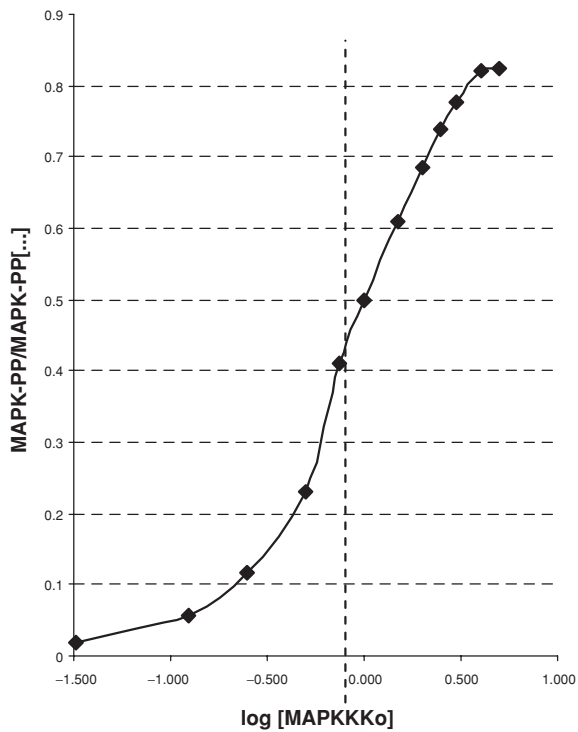
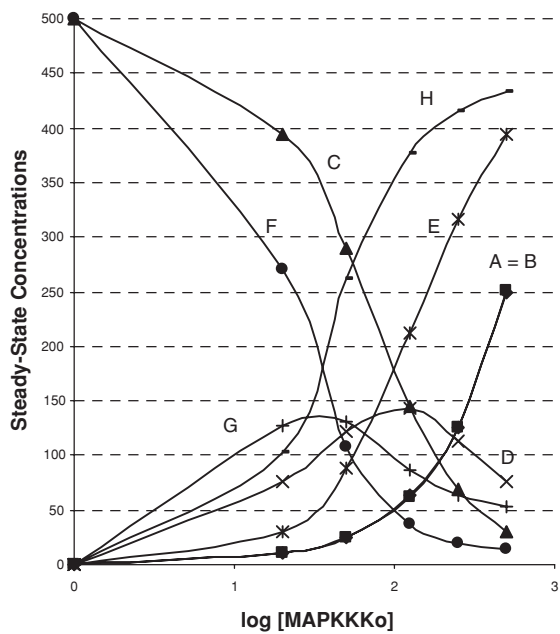


Figure 6.26. Influence of the MAPKK-protease propensity $P(E3)$ on the steady-state concentrations of the MAPK cascade species ($A = \text{MAPKKK}$, $B = \text{MAPKKK}^*$, $C = \text{MAPKK}$, $D = \text{MAPKK-P}$, $E = \text{MAPKK-PP}$, $F = \text{MAPK}$, $F = \text{MAPK-P}$, $H = \text{MAPK-PP}$). Enzyme propensities $P(E1) = P(E2) = P(E4) = 0.1$, substrate initial concentrations $[A_o] = 50$, $[C_o] = [F_o] = 500$.

concentrations of the substrates MAPKKK, MAPKK and MAPK, and constant propensity of the other three enzymes. We illustrate this type of pathway modeling in Figure 6.26 with the variation of the MAPKK-protease (E3 enzyme in Figure 6.1), which reverses the two-step reaction of MAPKK phosphorylation. It is shown that the concentration of the MAPKK- and MAPK-diphosphates (marked as E and H respectively) passes through a maximum near relatively low enzyme transition probability ($P \approx 0.02$). At the point of its maximum, the concentration of MAPK-PP reaches over 80% of its maximum, whereas that of MAPKK-PP is slightly over 50%. This shows the potential for a strong influence on the concentrations of the two diphosphates in the MAPK cascade by inhibiting the MAPKK-protease. In contrast, the level of steady-state concentrations of the two monophosphates



(marked by D and G in Figure 6.26) is not sensitive to the activity of the enzyme modeled, except for the extreme case of very strong inhibition ($P \rightarrow 0.001$).

Varying the concentration of MAPKKK at constant enzyme propensities

Another line of analysis is to study how the variations in the initial concentration of MAPKKK, the major downstream effector, affects the MAPK-cascade when keeping the enzyme activity constant. Figure 6.27a shows the dynamics of the concentrations of all substrates and products for a 25-fold range of $[MAPKKK_0]$ from 20 to 500 cells, constant enzyme propensities $P(E) = 0.1$, and constant initial concentrations of MAPKK and MAPK equal to 500 cells. In the semilogarithmic plot of Figure 6.27b, the MAPK-PP and MAPKK-PP ascending curves are sigmoids, as are the descending curves of the respective substrates MAPK and MAPKK, respectively. The sigmoidal pattern is also manifested in Figure 6.27b with the MAPK curve in predicted stimulus/response semilogarithmic plot, the input stimulus in which is expressed in multiples of the EC_{50} , the concentration of MAPKKK that produces 50% maximal response. This behavior is typical for allosteric enzymes disobeying Michaelis-Menten kinetics, and thus evidences for a cooperative effect of cascade enzymes. Our finding confirms the result of Huang and Ferrell [95], obtained by numerical solving of the differential rate equations, assuming the concentration of enzyme E1 as a basic variable.

Simultaneous variation of MAPKKK concentration and enzymes competence

The dynamics of the MAPK cascade signaling pathway can be analyzed in more details at the simultaneous variation of substrate concentrations and enzyme propensities. We varied the concentration of MAPKKK within a 25-fold range: 20, 50, 125, 250, 500 cells, keeping constant the

Figure 6.27. a) Semi-logarithmic plot of the steady-state concentrations of substrates and products dependence on the initial concentration of MAPKKK in MAPK cascade ($A = MAPKKK$, $B = MAPKKK^*$, $C = MAPKK$, $D = MAPKK-P$, $E = MAPKK-PP$, $F = MAPK$, $G = MAPK-P$, $H = MAPK-PP$). $P(E1) = P(E2) = P(E3) = P(E4) = 0.1$, $[C_0] = [F_0] = 500$; b) Predicted relative stimulus/response curve for MAPK-PP with input stimulus (the MAPKKK initial concentration) expressed in multiples of EC_{50} .

concentrations of MAPKK and MAPK at level 500 cells. The variable enzyme competence was studied within the 900-fold range from 0.001 to 0.9, keeping the other enzymes' competence at the 0.1 level. This type of CA modeling is better illustrated on contour plots showing the concentration profiles of products and substrates.

Varying $[MAPKKK_o]$ and enzyme E1 propensity did not provide surprising results because both variables increase the yield of reaction products. Yet, the CA models showed a pattern of dominance of the concentration of substrate A over the competence of the enzyme E₁. Thus, at $[MAPKKK_o] = 20$, the 45-fold increase in enzyme E1 activity from 0.02 to 0.90 results in only about 4.5 fold increase in the $MAPK \rightarrow MAPK-PP$ conversion ratio. In contrast, even at relatively low activity of E1 ($P = 0.02$), the increase of $[MAPKKK_o]$ from 20 to 250 (12.5-fold) increases the production of MAPK-PP 7-fold (from 50 to 350).

The variation of enzymes E2, E3, and E4 propensity produces contour plots with interesting features. This is illustrated for enzyme E3 in Figure 6.28, the contour lines in which show levels of constant steady-state concentration of MAPK-PP. The enzyme effect on the yield of MAPK-PP passes through a ridge at $P(E3) = 0.02$ to 0.1. The decrease in the MAPK-PP production is expected when E3 becomes more active, because this enzyme reduces the concentration of MAPKK-PP, which catalyzes the MAPK phosphorylation reaction. However, the increase in the MAPK-PP concentration within the $P(E3) = 0.001$ to 0.02 range of enzyme activity, for the broad range of $[MAPKKK] \geq 25$ cells, is a trend that hardly could be anticipated. One can assess from the contour plots analyses that suppressing the activity of E2, E3, and E4 enzymes to a level obtained at probability $P = 0.01$ -0.02 would enable reaching 80% of the maximum MAPK-PP concentration at relatively low concentrations of MAPKKK. A further maximization of the MAPK-PP production can result from a more favorable combination of the four enzyme propensities, a high one for enzyme 1 and low ones for enzymes E2, E3, and E4, as follows from the patterns described above. Conversely, a substantial inhibition of these enzymes (e. g., at $P \ll 0.02$) would minimize the MAPK-PP steady-state concentration.

Concluding our analysis, we would like to emphasize again that the CA modeling of the MAPK cascade pathway reported here was not aimed at exact reproducing of previous work [95]. Rather, it was aimed to demonstrate the potential of the cellular automata method to model basic patterns in pathways of interest and to indicate the ways to control the pathways

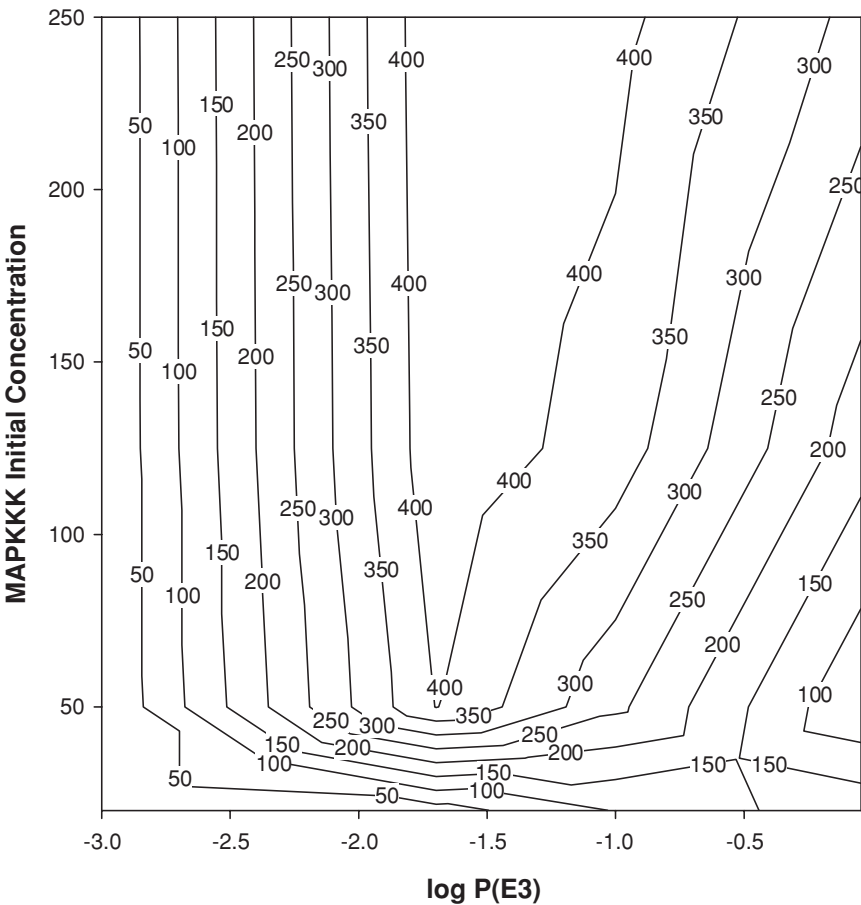


Figure 6.28. Contour plot of the MAPK-PP steady-state concentration at variable MAPKKK initial concentration and variable MAPKK-protease propensity. $P(E1) = P(E2) = P(E4) = 0.1$, $[MAPKK_o] = [MAPK_o] = 500$.

dynamics by selective enzyme inhibiting and concentration variations. Work is in progress to extend the methodology to large networks.

5. General Summary

The linkage between complex, dynamic systems and cellular automata is made quite clear in this monograph. The dynamic portrayal of many

phenomena has been shown to mirror reality in several important ways among the studies described. We aver that cellular automata belongs on the pantheon of methods of probing, modeling and even predicting events associated with complex systems, emerging phenomena and hierarchical patterns.

References

1. H. Gutfreund, *Kinetics for the Life Sciences: Receptors, Transmitters, and Catalysts*, Cambridge University, Cambridge, U.K. (1995).
2. M. A. Savageau, *Biochemical Systems Analysis*, Addison-Wesley, Reading, MA (1976).
3. D. L. Beveridge and W. L. Jorgensen, *Computer Simulation of Chemical and Biomolecular Systems*, Vol. 482, Annals of the New York Academy of Sciences, New York (1986).
4. L. A. Segel (ed.), *Biological Kinetics*, Cambridge University, Cambridge, UK (1991).
5. S. A. Levin (ed.), *Frontiers in Mathematical Biology*, Springer, Berlin (1994).
6. T. M. Witten, Computational Medicine, in *Encyclopedia of Computer Science* 4th edition, A. Ralston, E. D. Reilly, and D. Hemmendinger (eds.), Macmillan Reference, London and Grove's Dictionaries, New York (2000).
7. P. W. Anderson, K. J. Arrow, and D. Pines (eds.), *The Economy as an Evolving Complex System*, Santa Fe Institute, Studies in the Sciences of Complexity, Addison-Wesley, Redwood City, CA (1988).
8. M. S. Townend, *Mathematics in Sport*. Ellis Horwood, Chichester, UK (1984).
9. C. Renfrew and K. L. Cooke, *Transformations: Mathematical Approaches to Culture Change*, Academic Press, New York (1979).
10. A. S. Mikhailov and V. Calenbuhr, *From Cells to Societies: Models of Complex Coherent Interaction*, Springer, Berlin (2002).
11. W. Sulis and A. Combs, *Nonlinear Dynamics in Human Behavior*, World Scientific, Singapore (1996).
12. T. M. Witten, *Bull. Math. Biol.* **42**, 267-272 (1980).
13. T. M. Witten, *Bull. Math. Biol.* **44**, 572-584 (1982).
14. TM Witten, Computational Medicine, in: *Encyclopedia of Computer Science*, 4th ed., A. Ralston, E. D. Reilly, and D. Hemmendinger (eds.), Macmillan Reference, London, UK, and Grove's Dictionaries, New York (2000).
15. *National Science Foundation. Grand Challenges: High Performance Computing and Communications*. The FY 1992 US Research and Development Program (1992).
16. *National Science Foundation. Grand Challenges: High Performance Computing and Communications*. The FY 1993 US Research and Development Program (1993).
17. T. M. Witten, *Supercomputing Review* **3**, 34-40 (1990).
18. T. M. Witten, *Supercomputer*, **8**, 37-53 (1991).

19. T. M. Witten, *Future Generation Computer Systems* **10**, 223-232 (1994).
20. J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods*, Wiley, New York (1992).
21. A. R. Leach, *Molecular Modelling: Principles and Applications*, Longman, Harlow, UK (1996).
22. M. P. Allen and D. J. Tildesley, *Computer Simulation in Chemical Physics*, Kluwer Academic, Boston (1998).
23. D. J. Tildesley, The Molecular Dynamics Method, In: *ibid*, p. 23-47.
24. M. P. Allen, Introduction to Monte Carlo Simulations, in *Observation, Prediction and Simulation of Phase Transitions in Complex Fluids*, M. Baus, L. F. Rull, J.-P. Ryckaert (eds.), Kluwer Academic, Boston (1995) pp. 339-356.
25. W. L. Jorgensen, Monte Carlo Simulations for Liquids, in *Encyclopedia of Computational Chemistry*, P. V. Ragué Schleyer (ed.), Wiley, New York (1998) pp. 1754-1763.
26. W. L. Jorgensen and J. Tirado-Rives, *J. Phys. Chem.* **100**, 14508-14513 (1996).
27. R. K. Belew and Mmitchell (eds.), *Adaptive Individuals in Evolving Populations: Models and Algorithms*, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA (1996).
28. T. F. H. Allen and T. B. Starr, *Hierarchy: Perspectives for Ecological Complexity*, The University of Chicago, Chicago, IL (1982).
29. S. N. Salthe, *Evolving Hierarchical Systems: Their Structure and Representation*. Columbia University, New York (1985).
30. Y. Bar-Yam, *Dynamics of Complex Systems. Studies in Nonlinearity*. Perseus, Reading, MA (1997).
31. N. MacDonald, *Trees and Networks in Biological Systems*. Wiley, Chichester, UK (1983).
32. N. Rashevsky, *Organismic Sets*, Clowes and Sons, London, U.K. (1972).
33. R. Rosen, *Bull Math. Biol.* **20**, 245-260 (1958).
34. R. Rosen, *Bull Math. Biol.* **20**, 317-341(1958).
35. R. Rosen, *Bull Math. Biol.* **21**, 109-128 (1959).
36. R. Rosen, *Bull Math. Biol.* **27**, 11-14 (1965).
37. T. M. Witten, *Mech. Aging and Dev.*, **27**, 323-340 (1984).
38. E. H. Davidson, JP Rast, P.Oliveri, *Science* **295**, 1669-1678 (2002).
39. R. Satorras-Pastor, E. Smith, and R. V. Sole, *J.Theor. Biol.* **222**, 199-210 (2003).
40. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. -L. Barabasi, *Science* **297**, 1551-1555 (2002).
41. A. L. Barabasi and R. Albert , *Science* **286**, 509-512 (1999).
42. J. M. Montoya and R. V. Sole, *J. Theor. Biol.* **214**, 405-412 (2002).
43. R. Albert and A. -L. Barabasi, *Rev. Mod.Phys.* **74**, 47-97 (2002).
44. J. -P. Eckmann, E. Moses, and D. Sergi, Dialog in E-Mail traffic. E-print archives, <http://xxx.aps.org/arXiv:cond-mat/0304433> (2003).

45. D. Bonchev, Overall Connectivity and Molecular Complexity, in *Topological Indices and Related Descriptors*, J. Devillers and A. T. Balaban (eds.), Gordon and Breach, Reading, UK (1999) pp. 361-401.
46. D. Bonchev, The Wiener Number. Some Applications and New Developments, in *Topology in Chemistry. Discrete Mathematics of Molecules*, D. H. Rouvray and R. B. King (eds.), Horwood, Chichester, UK (2002) pp. 58-88.
47. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. -L. Barabasi, *Science* **297**, 1551-1555 (2002).
48. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824-827 (2002).
49. M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York, NY, Simon and Schuster (1992).
50. P. Decker, Spatial, Chiral, and Temporal Self-Organization Through Bifurcation in "Bioids," Open Systems Capable of a Generalized Darwinian Evolution, in *Bifurcation Theory and Applications in Scientific Disciplines*, O. Gurel, and O. E. Rossler (eds.), Annals of the New York Academy of Science, New York, NY (1979) Vol. 316.
51. L. F. Olsen and H. Degn, *Nature* **267**, 177 (1977).
52. E. E. Selkov, *Eur. J. Biochem.* **4**, 79 (1967).
53. B. Hess and T. Plesser, Temporal and Spatial Order in Biochemical Systems, in *Bifurcation Theory and Applications in Scientific Disciplines*, O. Gurel and O. E. Rossler (eds.), Annals of the New York Academy of Science, New York, NY (1979) Vol. 316.
54. K. R. Sharma and R. M. Noyes, *J. Amer. Chem. Soc.* **98**, 4345 (1976).
55. G. Nicolis and I. Prigogine, *Exploring Complexity: An Introduction*. New York, NY, Freeman (1989).
56. U. S. Bhalla and R. Iyengar, *Science* **283**, 381-387 (1999).
57. L. B. Kier, B. Testa and P. A. Carrupt, *Med. Res. Rev.* **17**, 303-326 (1997).
58. B. Testa, and L. B. Kier, *Drug Res.* **30**, 1-14 (1997).
59. H. G. Wells, J. S. Huxley, and G. P. Wells, *The Science of Life: The Literary Guild*, New York (1934) p. 1475.
60. S. M. Ulam, *Proc. Int. Congr. Math.* **2**, 264 (1952).
61. S. M. Ulam, *Adventures of a Mathematician*, Charles Scribner's Sons, New York (1976).
62. J. von Neumann, *Theory of Self-Replicating Automata*, A. Burks (ed.), Univ. of Illinois Press, Urbana (1966).
63. K. Zuse, *Int. J. Theoret. Phys.* **21**, 589-600 (1982).
64. T. Toffoli and N. Margolus, *Cellular Automata Machines*, MIT Press, Cambridge, MA (1987).
65. M. Schroeder, *Fractals, Chaos, Power Laws*, W. H. Freeman, New York (1991) p. 371.
66. G. Y. Vichniac, *Physica D*, **10**, 96-116 (1984).
67. T. Toffoli, *Physica D*, **10**, 117-127 (1984).
68. S. Wolfram, *Rev. Mod. Phys.* **55**, 601-644 (1983).
69. L. B. Kier and C. K. Cheng, *J. Chem. Inf. Comp. Sci.* **34**, 1334-1337 (1994).

70. L. B. Kier and C. K. Cheng, *Pharm. Res.* **12**, 1521-1525 (1995).
71. L. B. Kier and C. K. Cheng, *J. Math. Chem.*, **21**, 71-81 (1997).
72. L. B. Kier, C. K. Cheng, B. Testa, and P. A. Carrupt, *Pharm. Res.* **12**, 615-620 (1995).
73. C. K. Cheng and L. B. Kier, *J. Chem. Inf. & Comput. Sci.* **35**, 1054-1059 (1995).
74. L. B. Kier and C. K. Cheng, in *Lipophilicity in Drug Research*, Pliska, Testa, and Waterbeemd (eds.), VCH (1996).
75. L. B. Kier, C. K. Cheng, B. Testa and P. A. Carrupt, *Pharm. Res.* **13**, 1419-1422 (1996).
76. L. B. Kier, C. K. Cheng, B. Testa and P. A. Carrupt, *J. Pharm. Sci.* **86**, 774-779 (1997).
77. L. B. Kier and C. K. Cheng, *J. Theor. Biol.* **186**, 75-80 (1997).
78. L. B. Kier, C. K. Cheng, M. Tute, and P. G. Seybold, *J. Chem. Inf. Comp. Sci.* **38**, 271-280 (1998).
79. L. B. Kier, C. K. Cheng and B. Testa, *J. Chem. Inf. Comp. Sci.* **39**, 326-332 (1999).
80. L. B. Kier, C. K. Cheng and B. Testa, *Fut. Gen. Comput. Sys.* **16**, 273-289 (1999).
81. L. B. Kier, C. K. Cheng and P. Seybold, *SAR QSAR Environ. Res.* **11**, 79-102 (2000).
82. L. B. Kier, C. K. Cheng and P. Seybold, *Rev. Comp. Chem.* **17**, 205-254 (2001).
83. L. B. Kier and C. K. Cheng, *J. Chem. Inf. Comp. Sci.* **34**, 647-652 (1994).
84. L. B. Kier, Cheng and B. Testa, *J. Theoret. Biol.* **214**, 415-426 (2002).
86. G. R. Welch, *Biosystems* **38**, 147-153 (1996).
87. G. Adam and M. Delbruck, in *Structural Chemistry and Molecular Biology*, A. Rich and N. Davidson (eds.), Freeman, San Francisco (1968).
88. M. Eigen, in *Quantum Statistical Mechanics in the Natural Sciences*, S. L. Mintz and S. M. Widemeyer (eds.), Plenum, New York (1974).
89. G. R. Welch, *J. Theoret. Biol.* **68**, 267-271 (1977).
90. L. B. Kier and B. Testa, *Complexity*, **1**, 37-41 (1996).
91. C. H. Waddington, *The Strategy of the Genes*, George Allen & Urwin Ltd., London, UK (1957).
92. L. B. Kier, *AANA Journal* **71**, 422-426 (2003).
93. D. Bonchev, Complexity of Protein-Protein Interaction Networks, Complexes and Pathways, in *Handbook of Proteomics Methods*, M. Conn (ed.), Humana, New York, (2003), pp. 451-462.
94. L. B. Kier, D. Bonchev and G. A. Buck. *Chem. & Biodiv.*, **2**, 233-243 (2005).
95. C. Y. F. Huang and J. E. Ferrell, *Proc. Natl. Acad. Sci. USA* **93**, 10078-10083 (1996).
96. L. B. Kier, C. K. Cheng, and B. Testa, *J. Molec. Graphics* **14**, 227-231 (1996).
97. L. B. Kier and C. K. Cheng, *J. Molec. Graphics* **18**, 29-32 (2000).

Chapter 7

THE COMPLEX NATURE OF ECODYNAMICS

Robert E. Ulanowicz

University of Maryland, Center for Environmental Science, Chesapeake Biological Laboratory,
P.O. Box 38, Solomons, MD 20688 USA

1. Introduction

The specific nature of ecodynamics is rarely discussed. The tacit assumption usually is that ecosystems behave like most of the rest of nature—according to the laws of conservation of material and energy and obeying a set of determinate dynamical laws—much like those that govern planetary motions. Furthermore, most continue to assume that the same mathematics that so aptly quantifies the physical world is sufficient to illumine ecodynamics. Unfortunately, precious few results have been achieved to date under this agenda, but ecologists continue nonetheless to exhibit what Cohen has somewhat facetiously referred to as “physics envy”[1]. It would be unfair to accuse only ecologists of outmoded thinking. Much of what has appeared in the literature under the rubric of “complexity theory” has proceeded along those same lines: Complexity is considered but an epiphenomenon of scale. Complicated behavior is still thought by many to be the result of very simple interactions at smaller levels that ramify at larger scales to yield strange and manifold behaviors, such as are described by the use of fractal theory. Such “thin” complexity remains business as usual [2], only with a few non-linear wrinkles thrown in. Whence, Sally Goerner [3] portrays most of Complexity Theory as “21st Century science built upon 19th Century foundations.”

The opinion is beginning to emerge, however, that complexity may require an essentially different way of viewing how nature works (e.g., Casti [4], Kay and Schneider [5], Salthe [6], Rosen [7], Mikulecky [8].) Fortunately, those who see complexity in this new guise are usually not reticent about expressing their dissent. The mathematician, John Casti [9], for example, builds upon a children’s story, “Little Bear”, by Else Minarik [10]

wherein the principal character tries to get to the moon by climbing a tree. Casti contends that using the conventional methods of physics to improve one's understanding of complex systems is akin merely to climbing a taller tree. He cites how, when complex systems are approached using conventional tools, they almost invariably give qualitatively *contradictory* prognoses (and not ones that are merely quantitatively inaccurate.)

Of course, none of the dissenters is contending that ecosystems, or any other living systems for that matter, violate the laws of conservation of material and energy. Unlike the realms of relativity theory and quantum physics, there is nothing about ecodynamics that would place those dogmas into question. In fact, one might even argue that ecosystems obey these dicta too well. Matter and energy in ecosystems usually return in very quick order to being almost in balance and thereby render these laws of little assistance in revealing what the system is doing over the longer term.

The evolutionary theorist might respond that what is missing from the methods of physics is the historical approach initiated by Darwin in his theory of descent by natural selection. They would indeed be correct in crediting Darwin with the introduction of history into science. In fact, Darwin had a healthy respect for the complex nature of evolution [11]. Unfortunately, most of Darwin's nuances were ignored during the formulation of the "Grand Synthesis" by Fisher and Sewall Wright, who precipitated what now is called "neo-Darwinism". In this contemporary wisdom, it is the information encoded in the genome of organisms that directs the behavior of living entities, and, by aggregative induction, those of the whole ecosystem. Those who view evolution in such simplistic fashion conveniently ignore the problem of how it forces the observer constantly to switch back and forth in almost schizoid manner between the contingencies in genomic reproduction and the presumably lawful behavior of the resulting phenome in its environment. It will suffice here, however, simply to note that the neo-Darwinian approach does not seem to be applicable to ecosystems. As Guenther Stent [12] so aptly put it,

"Consider the establishment of ecological communities upon colonization of islands or the growth of secondary forests. Both of these examples are regular phenomena in the sense that a more or less predictable ecological structure arises via a stereotypic pattern of intermediate steps, in which the relative abundances of various types of flora and fauna follow a well-defined sequence. The regularity of these phenomena is obviously not the consequence of an ecological program encoded in the genomes of

the participating taxa [13].” Neither has elucidating the ontogenetic mapping from genome to phenome been any raging success. Efforts by Sidney Brenner et al. (ibid.) to identify the connections, together with recent results from the Human Genome Project [14] reveal that the full mapping is very likely a chimera. As Brenner bravely suggested, it becomes necessary to “try to discover the principles of organization, how lots of things are put together in the same place” [13]. It happens that Brenner’s is a relational task that is tailor-made for the ecologist.

The reader would be justified in asking how Brenner’s suggested approach differs from how problems are normally posed in physics. In that epitome of the hard sciences problems are usually parsed into what are called the field equations and the boundary conditions. Although one usually wishes to study a phenomenon over a given domain (field) of space and time, it is assumed that one needn’t measure the magnitude of the phenomenon at all points of the field. Rather, a particular law expresses in a very compact way how the given attribute will vary from point to point within the field. It becomes necessary, therefore, to specify only the magnitude of the phenomenon at the peripheries of the field (and at the initial time.) Important here is the fact that all the known laws of physics are entirely symmetrical with respect to time [15]. They cannot impart any asymmetry to the field with which the observer may distinguish it from an adirectional background [16]. In other words, uniqueness and direction enter the system *only* via the imposed boundary constraints.

Turning now to complex biotic systems that can be parsed into a number of essential components, these elements respond to some degree (like physical systems) to constraints arising *outside* the ensemble. Those exogenous constraints, however, are insufficient to determine the behavior of the component, because the components themselves interact with one another. That is, each component is constrained by, and in its own turn constrains, other compartments. (This is most unlike the systems that Boltzmann or Fischer studied, which were collections of many *non-interacting* elements.) In ecosystems and other biotic communities the boundary constraints on any element arise in part *within* the system itself as engendered by other proximate elements.

If one represents the constraint exerted by one compartment upon another as a directed link, then these links are wont to combine with one another into chains of constraints, which in some instances can fold back upon themselves and form cycles. When this latter circumstance occurs, the

participating elements exert a degree of constraint upon themselves that traces back to no external source. Robert Rosen [7] defined organisms as systems that were self-entailing with respect to efficient causes. That is, the agencies behind repair, growth and metabolism are all elicited by each other and do not derive from any external source. Similarly, closed cycles of constraints set the stage for some internal (partially) autonomous control of biotic systems.

The controlling nexus of ecodynamics now becomes clearer. It is not the field equations of conservation of mass and energy that are of greatest interest. These are nearly satisfied in relatively short order. Nor are energy-based constraints (e.g., ecosystems develop so as to store the maximal amount of exergy possible [17] alone sufficient to dictate the final outcome [although such global constraints most probably do affect the endpoint.]) It appears as concerns transitional ecodynamics that the paramount focus should be upon the interactions among the (mostly hidden) internal constraints, which change more slowly with time. That is, control of ecodynamics appears to be relational in nature—how much any change in one constraint affects others with which it is linked. As Stent suggested, changes in genomic constraints remain hidden in this perspective; and, furthermore, there is no obvious reason to suspect that they are cryptically directing matters. In fact, it could even be argued that the internally closed loops of constraints serve, over the longer run, to sift among genetic variations and to select *for* those that accord better with their own actions, as will be developed below.

2. Measuring the Effects of Incorporated Constraints

Such theorizing may be all for the good, but science requires measurement and quantification as well. Any well-posed theory must have the potential to become operational. Herein lies a possible stumbling-block, because there is simply no hope of making explicit, much less measuring, every item of internal constraint in any living system (the Human Genome Project notably *notwithstanding*.) But this quandary is not unknown to those familiar with thermodynamics and statistical mechanics. There one is confronted with effects stemming from an unmanageable number of atomic entities, and it is impossible to follow the actions of each actor in detail. So, rather than attempting to quantify the trajectories of each individual actor, physical attributes of the entire (macroscopic) ensemble are measured.

Whole system properties, such as pressure, temperature and volume, are assumed to be common attributes upon which have been implicitly written the contributions of each microscopic event.

This same stratagem can also be applied to ecosystems. One begins by acknowledging the importance of each internal constraint, such as prey escape tactics, mating displays, visual cues, etc. The focus, however, is upon the measurement (or at least estimation) of more aggregated processes, such as how much material and/or energy passes from one system element to another over a given interval of time. All such estimated transfers can then be arrayed as a network of ecosystem material and/or energy linkages—diagrams of “who eats whom, and at what rates?” This “brutish” description [18] of ecosystem behavior at first glance appears to ignore most of what interests biologists and what imparts pattern to the ecosystem, but in the spirit of thermodynamics, those vital elements are assumed to write their effects upon this “macroscopic” quantification of ecosystem behavior. Change any one of the hidden constraints, and its consequence(s) will be observed, at least incrementally, upon the network of system flows [19].

Just as the aggregated effects of individual agents are captured by the macroscopic variables of thermodynamics, so does an ecosystem flow network embody all the consequences of the hidden constraints. It remains, however, to quantify the effects of existing embodied constraints upon this pattern over and against other confounding factors that may affect the network structure of the system. Before doing so, however, it is necessary first to avoid the significant temptation to assume that closed circuits of concatenated constraints are merely another mechanical agency. With ecosystem networks one is dealing instead with an essentially different dynamics, which is made apparent by two significant points: (1) Constraints in living systems are not rigidly mechanical in nature, but incorporate singular contingencies in a necessary but limited way. (2) Cyclical relationships among some constraints, by virtue of the singular events they incorporate, give rise to categorically non-mechanical agencies.

3. Ecosystems and Contingency

That living systems are not fully constrained, i.e., that they retain sufficient flexibility to adapt to changing circumstances, is (along with self-entailment) a necessary attribute of living systems. It should become

apparent, furthermore, that the tension between constraint and its complement, flexibility is probably easier to discern in ecosystems than in organisms, where the constraints are more prevalent and rigid; for every ecologist is acutely aware of the significant role that the aleatoric plays in ecology. Chance events occur everywhere in ecosystems. Stochasticity is hardly unique to ecology, however, and the entire discipline of probability theory has evolved to cope with contingencies. Unfortunately, few stop to consider the tacit assumptions made when invoking probability theory—namely that chance events are always simple, generic and recurrent. If an event is not simple, or if it occurs only once for all time (is truly singular), then the mathematics of probabilities will not apply.

It may surprise some, therefore, to learn that ecosystems appear to be rife with singular events [20]. To see why, it helps to recall an argument formulated by physicist Walter Elsasser [21]. Elsasser sought to delimit what he called an “enormous” number. By this he was referring to numbers so large that they should be excluded from physical consideration, because they greatly exceed the number of physical events that possibly could have occurred since the Big Bang. To estimate a threshold for enormous numbers Elsasser reckoned the number of simple protons in the known universe to be about 10^{85} . He then noted as how the number of nanoseconds that have transpired since the beginning of the universe has been about 10^{25} . Hence, a rough estimate of the upper limit on the number of conceivable events that could have occurred in the physical world is about 10^{110} . Any number of possibilities much larger than this value simply loses any meaning with respect to physical reality.

Anyone familiar with combinatorics immediately will realize that it doesn't take very many distinguishable elements or processes before the number of their possible configurations becomes enormous. One doesn't need Avagadro's Number of particles (10^{23}) to produce combinations in excess of 10^{110} —a system with merely 80 or so distinct components will suffice. In probabilistic terms, any event randomly comprised of more than 80 distinct elements is virtually certain never have occurred earlier in the history of the universe. Such a constellation is unique over all time past. It follows, then, that in ecosystems with hundreds or thousands of distinguishable organisms, one must reckon not just with occasional unique events, but rather with a legion of them. Unique, singular events are occurring *all the time, everywhere!* In the face of this reality, one must abandon

any hope of determinism as a universal characteristic of natural systems, and it becomes difficult as well to conceive of living systems as reversible.

Despite the challenge that rampant singularities pose for the Baconian pursuit of science, a degree of regularity can still be observed in such ecological phenomena as succession. The question then arises as to the origins and maintenance of such order? Unfortunately, the conventional evolutionary narrative is constantly switching back and forth between the realms of strict determinism and pure stochasticity, as if no middle ground might exist. In referring to this regrettable situation, Karl Popper [22] remarked that it still remains for science to achieve a truly “evolutionary theory of knowledge”, and one will not be forthcoming until fundamental attitudes toward the nature of causality have been reconsidered. True reconciliation, Popper suggested, lies in envisioning an intermediate to stochasticity and determinism. To meet this challenge, he proposed a generalization of the Newtonian conception of “force”. Forces, he posited, are idealizations that exist as such only in perfect isolation. The objective of experimentation is to approximate isolation from interfering factors as best possible. In the real world, however, where components are loosely, but definitely coupled, one should speak rather of “propensities”. A propensity is the tendency for a certain event to occur in a particular context. It is related to, but not identical to, conditional probabilities.

Consider, for example, the hypothetical “table of events” depicted in Table 7.1, which arrays five possible outcomes, b_1, b_2, b_3, b_4, b_5 , according to four possible eliciting causes, a_1, a_2, a_3 , and a_4 . For example, the outcomes might be several types of cancer, such as those affecting the lung, stomach, pancreas or kidney, while the potential causes might represent various forms of behavior, such as running, smoking, eating fats, etc. In an

Table 7.1. Frequency table of the hypothetical number of joint occurrences that four “causes” ($a_1 \dots a_4$) were followed by five “effects” ($b_1 \dots b_5$)

	b1	b2	b3	b4	b5	Sum
a1	40	193	16	11	9	269
a2	18	7	0	27	175	227
a3	104	0	38	118	3	263
a4	4	6	161	20	50	241
Sum	166	206	215	176	237	1000

Table 7.2. Frequency table as in Table 7.1, except that care was taken to isolate causes from each other.

	b1	b2	b3	b4	b5	Sum
a1	0	269	0	0	0	269
a2	0	0	0	0	227	227
a3	263	0	0	0	0	263
a4	0	0	241	0	0	241
Sum	263	269	241	0	227	1000

ecological context, the b 's might represent predation by predator j , while the a 's could represent donations of material or energy by host i .

One notices from the table that whenever condition a_1 prevails, there is a propensity for b_2 to occur. Whenever a_2 prevails, b_5 is the most likely outcome. The situation is a bit more ambiguous when a_3 prevails, but b_1 and b_4 are more likely to occur in that situation, etc. Events that occur with smaller frequencies, e.g., $[a_1, b_3]$ or $[a_1, b_4]$ result from what Popper calls “interferences”.

One now considers how the table of events might appear, were it possible to completely isolate phenomena, that is, were it possible to impose further constraints that would keep both other propensities and the arbitrary effects of the surroundings from influencing a given particular constraint? Probably, the result would look something like Table 7.2, where every time a_1 occurs, it is followed by b_2 ; every time a_2 appears, it is followed by b_5 , etc. That is, under isolation, propensities degenerate into mechanical-like forces. It is interesting to note that b_4 never appears under any of the isolated circumstances. Presumably, it arose purely as a result of interferences among propensities. Thus, the propensity for b_4 to occur whenever a_3 happens is an illustration of Popper’s assertion that propensities, unlike forces, never occur in isolation, nor are they inherent in any object. They always arise out of a context, which invariably includes other propensities.

In light of the above discussion, one could view Popper’s propensity as a constraint that is unable to perform unerringly in the face of confounding contingencies. Propensity encompasses under a single rubric the entire range of phenomena from singular events, through common chance, all the way to law-like behavior. One notices further that the transition depicted from Table 7.1 to Table 7.2 was accompanied by the addition of constraints, and it is the appearance of such progressive constraints that

one implies when one invokes the term “development”. Returning then to the second question at the end of Section 2, one now asks how the incorporation of the aleatoric moves the ensuing dynamics out of the realm of the purely mechanical?

4. Autocatalysis and Non-Mechanical Behavior

It was mentioned above how constraints can be concatenated and in some cases join back upon themselves (form cyclical configurations.) It has not yet been mentioned that constraints of one process upon another can be either excitatory (+) or inhibitory (−). It happens that the configuration of reciprocal excitation, or mutualism (+, +) can exhibit some very interesting behaviors that, in connection with aleatoric events, qualify its action as a non-mechanical causal agency. Investigators such as Manfred Eigen [23], Hermann Haken [24], Maturana and Varela [25], Stuart Kauffman [26] and Donald DeAngelis [27] all have contributed to the growing consensus that some form of positive feedback is responsible for most of the order one perceives in organic systems. It is useful now to focus attention upon a particular form of positive feedback, namely, autocatalysis.

Autocatalysis is positive feedback across multiple links wherein the effect of each and every link in the feedback loop upon the next remains positive. To be more precise, the reader’s attention is drawn to the three-component interaction depicted in Figure 7.1. It is assumed that the action of process A has a propensity to augment a second process B. It must be emphasized that the use of the word “propensity” implies that the excitatory constraint that A exerts upon B is not wholly obligatory, or mechanical. Rather, when process A increases in magnitude, most (but not all) of the

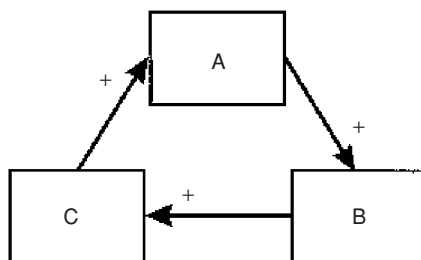
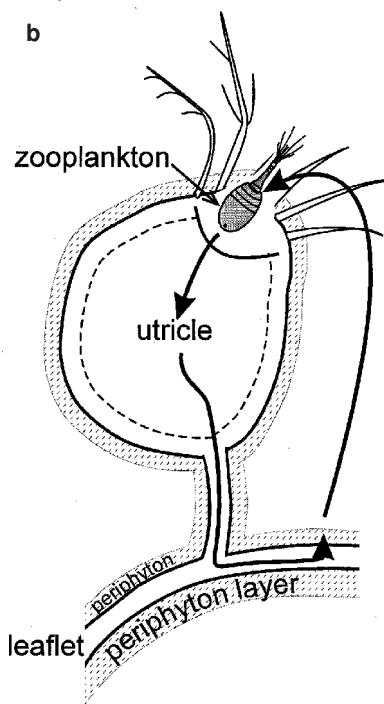
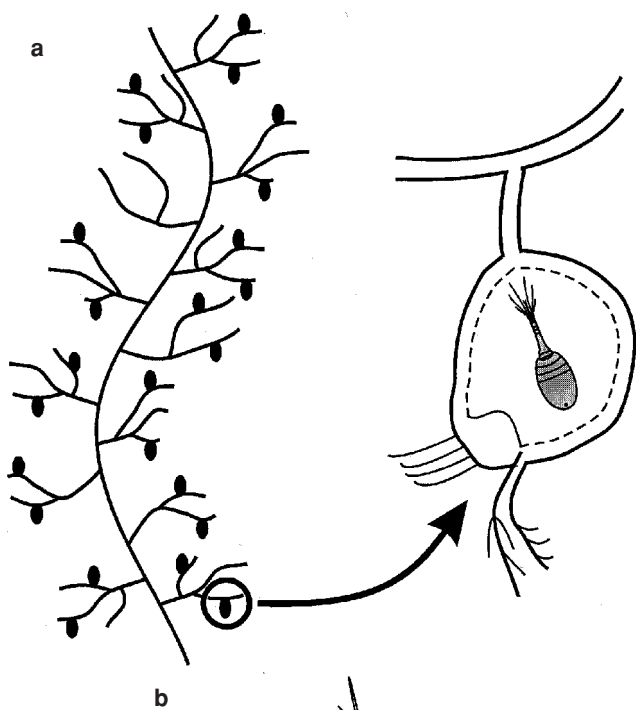


Figure 7.1. Schematic of a hypothetical 3-component autocatalytic cycle.



time, B also will increase. B tends to accelerate C in similar fashion, and C has the same effect upon A.

An ecological example of autocatalysis is the community that centers around the aquatic macrophyte, *Utricularia*, or bladderworts [28]. All members of the genus *Utricularia* are carnivorous plants. Scattered along its feather-like stems and leaves are small bladders, called utricles (Figure 7.2a). Each utricle has a few hair-like triggers at its terminal end, which, when touched by a feeding zooplankton opens the end of the bladder and the animal is sucked into the utricle by a negative osmotic pressure that the plant had maintained inside the bladder. In the field *Utricularia* plants always support a film of algal growth known as periphyton (Figure 7.2b). This periphyton in turn serves as food for any number of species of small zooplankton. The catalytic cycle is completed when the *Utricularia* captures and absorbs many of the zooplankton.

Autocatalysis among propensities gives rise to at least eight system attributes, which, taken as a whole, comprise a distinctly non-mechanical dynamic. One first notes that by the definition adopted here, autocatalysis is explicitly *growth-enhancing*. Furthermore, autocatalysis exists as a relational or *formal* structure of kinetic elements. Far more interesting is the observation alluded to earlier that autocatalysis is capable of exerting *selection* pressure upon all characteristics of its ever-changing constituents. To see this, one assumes that some small chance alteration occurs spontaneously in process B. If that change either makes B more sensitive to A or a more effective catalyst of C, then the change will receive enhanced stimulus from A. Conversely, if the change in B either makes it less sensitive to the effects of A or a weaker catalyst of C, then that change will likely receive diminished support from A. It is seen that that such selection works on the processes or mechanisms as well as on the elements themselves. Hence, any effort to simulate development in terms of a fixed set of mechanisms is doomed ultimately to fail.

It should be noted in particular that any change in B is likely to involve a change in the amounts of material and energy that flow to sustain B.

Figure 7.2. (a) Sketch of a typical “leaf” of *Utricularia floridana*, with detail of the interior of a utricle containing a captured invertebrate. (b) Schematic of the autocatalytic loop in the *Utricularia* system. Macrophyte provides necessary surface upon which periphyton (striped area) can grow. Zooplankton consumes periphyton, and is itself trapped in bladder and absorbed in turn by the *Utricularia*.

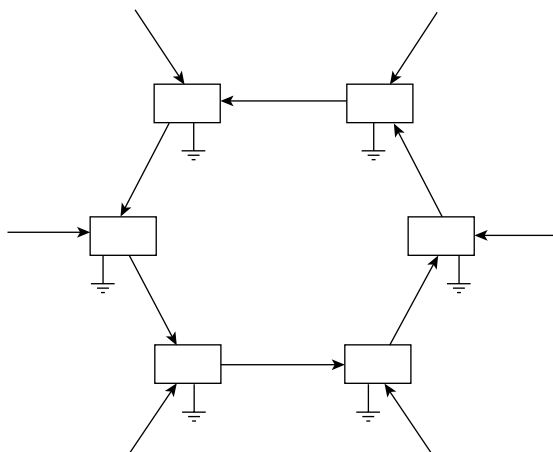


Figure 7.3. Centripetal action as engendered by autocatalysis.

Whence, as a corollary of selection pressure, one perceives a tendency to reward and support those changes that bring ever more resources into B. As this circumstance pertains to all the other members of the feedback loop as well, any autocatalytic cycle becomes the center of a *centripetal* vortex, pulling as many resources as possible into its domain (Figure 7.3.).

It follows that, whenever two or more autocatalytic loops draw from the same pool of resources, autocatalysis will *induce competition*. In particular, one notices that whenever two loops partially overlap, the outcome could be the exclusion of one of the loops. In Figure 7.4, for example, element D is assumed to appear spontaneously in conjunction with A and C. If D is more sensitive to A and/or a better catalyst of C, then there is a likelihood that the ensuing dynamics will so favor D over B, that B will either fade into the background or disappear altogether. That is, selection pressure and centripetality can guide the replacement of elements. Of course, if B can be replaced by D, there remains no reason why C cannot be replaced by E or A by F, so that the cycle A, B, C could eventually transform into D, E, F. One concludes that the characteristic lifetime of the autocatalytic form usually exceeds that of most of its constituents. This is not as strange as it may first seem. With the exception of neurons, virtually none of the cells that compose the human body persist longer than seven years. Very few of the atoms in it at a given time were present eighteen months earlier. Yet if

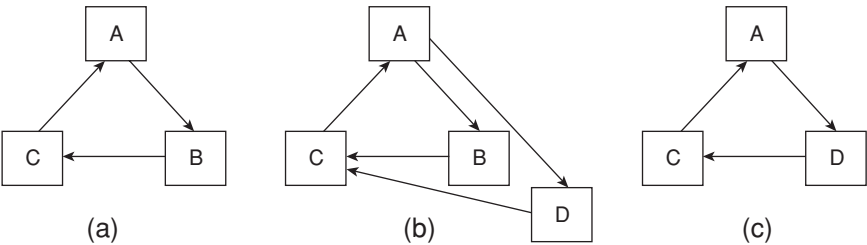


Figure 7.4. (a) Original configuration. (b) Competition between component B and a new component D, which is either more sensitive to catalysis by A or a better catalyst of C. (c) B is replaced by D, and the loop section A-B-C by that of A-D-C.

a mother were to see her offspring for the first time in ten years, chances are she would recognize him/her immediately.

Autocatalytic selection pressure and the competition it engenders define a preferred direction for the system—that of ever-more effective autocatalysis. In the terminology of physics, autocatalysis, predicated as it is upon eliciting internal constraints, each of which can be asymmetric, is therefore itself *symmetry-breaking*. One should not confuse this rudimentary directionality with full-blown teleology, however. It is not necessary, for example, that there exists a pre-ordained endpoint towards which the system strives. The direction of the system at any one instant is defined by its state at that time, and the state changes as the system develops. Perhaps the simple Greek term “*telos*” connotes better this weaker form of directionality and distinguishes it from the far rarer and more complex behavior known as teleology.

Taken together, selection pressure, centripetality and a longer characteristic lifetime all speak to the existence of a degree of *autonomy* of the larger structure from its constituents. Again, it must be stressed that attempts at reducing the workings of the system to the properties of its composite elements will remain futile over the long run.

In epistemological terms, the dynamics just described can be considered *emergent*. In Figure 7.5, for example, if one should consider only those elements in the lower right-hand corner (as enclosed by the solid line), then one can identify an initial cause and a final effect. If, however, one expands the scope of observation to include a full autocatalytic cycle of processes (as enclosed by the dotted line), then the system properties just described appear to emerge spontaneously.

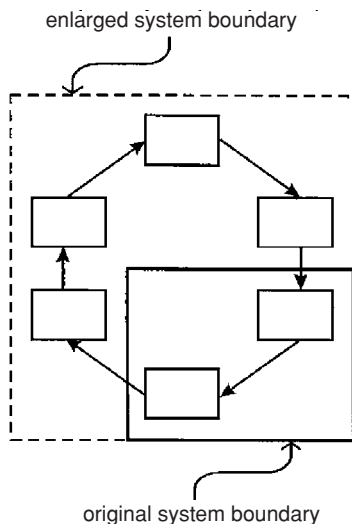


Figure 7.5. Two hierarchical views of an autocatalytic loop. The original perspective (solid line) includes only part of the loop, which therefore appears to function quite mechanically. A broader vision encompasses the entire loop, and with it several non-mechanical attributes.

5. Causality Reconsidered

Autocatalysis is thus seen to behave in ways quite uncharacteristic of machines. It is important also to note that the causal agency of autocatalysis appears in a form that is foreign to conventional mechanical analysis. In particular, the selection pressure that arises from autocatalysis acts from higher scales *downwards*. Conventional wisdom allows only influences originating at lower realms of time and space to exert their effects at larger and longer scales (reductionism.) This convention is a legacy of the Newtonian worldview and the ensuing Enlightenment. Prior to Newton, however, the prevailing view on natural causalities had been formulated by Aristotle, who explicitly recognized the existence of downward causation.

Aristotle identified four categories of cause: (1) Material, (2) Efficient (or mechanical), (3) Formal and (4) Final. An effective, albeit unsavory, example of an event wherein all four causes are at work is a military battle. The swords, guns, rockets and other weapons comprise the material causes of the battle [29]. The soldiers who use those weapons to inflict unspeakable harm on each other become the efficient agents. The topography of the

battlefield and the changing positions of the troops on the battlefield with respect to each other and with respect to natural factors, such as sun angle and wind, constitute the formal cause. Final cause originates mostly beyond the battlefield and consists of the social, economic and political factors that brought the armies to face each other.

Encouraged by the simplicity of Newton's *Principia* and perhaps influenced by the politics of the time, early Enlightenment thinkers acted decisively to excise formal and final causalities from all scientific description. Contemporary thinkers, such as the late Robert Rosen [30], are urging a reconsideration of whether these discarded categories might not serve the interpretation of complex phenomena. Indeed, there appear to be especial reasons why Aristotle's scheme provides a more satisfactory description of ecological dynamics, and those reasons center around the observation that efficient, formal and final causes are hierarchically ordered—as becomes obvious when one notices that the domains of influence by soldier, officer and prime minister extend over progressively larger and longer scales. It becomes apparent that autocatalytic loops of constraints are acting in the sense of formal agency (much like the ever-shifting juxtaposition of troops on the battlefield), selecting *for* changes among the participating ecosystem components.

The Achilles heel of Newtonian-like dynamics is that it cannot in general accommodate true chance or indeterminacy (whence the “schizophrenia” in contemporary biology.) Should a truly chance event happen at any level of a strictly mechanical hierarchy, all order at higher levels would be doomed eventually to unravel. The Aristotelian hierarchy, however, is far more accommodating of chance. Any spontaneous efficient agency at any hierarchical level would be subject to selection pressures from formal autocatalytic configurations above. These configurations in turn experience selection from still larger constellations in the guise of final cause, etc. One may conclude, thereby, that the influence of most irregularities remains circumscribed. Unless the larger structure is particularly vulnerable to a certain type of perturbation (and this happens relatively rarely), the effects of most perturbations are quickly damped.

One might even generalize from this “finite radius of effect” that the very laws of nature might be considered to have finite, rather than universal, domain (Allen and Starr [31], Salthe [32]). That is, each law is formulated within a particular domain of time and space. The farther removed an observed event is from that domain, the weaker becomes the explanatory

power of that law, because chance occurrences and selection pressures arise among the intervening scales to interfere with the given effect. To the ecologist, therefore, the world appears as granular, rather than universal.

6. Quantifying Constraint in Ecosystems

With these considerations on contingency, autocatalysis and causality in mind, one may now embark upon quantifying the overall degree of constraint in an ecosystem as manifested by its network of material/energy flows. Two major facets pertaining to the action of autocatalysis are relevant here: (a) Autocatalysis serves to increase the activities of all its constituents, and (b) it prunes the network of interactions so that those links that most effectively participate in autocatalysis become dominant. Schematically this transition is depicted in Figure 7.6. The upper figure represents a

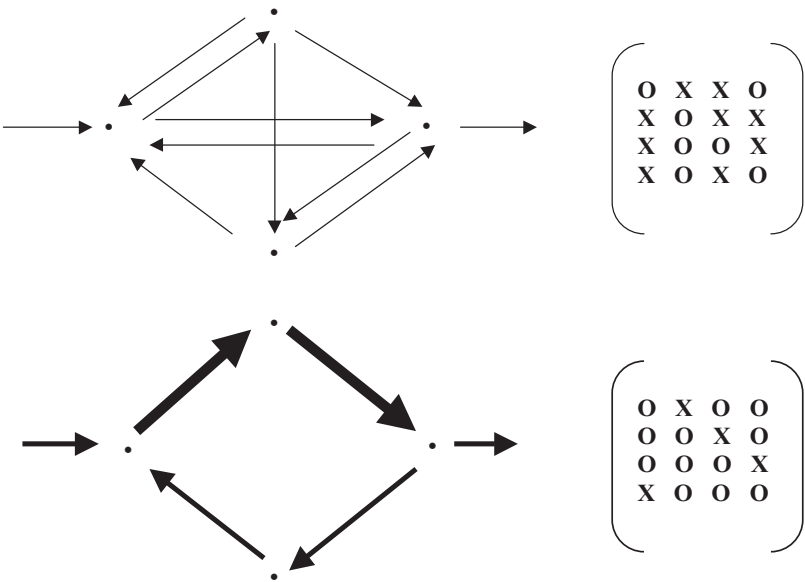


Figure 7.6. Schematic representation of the major effects that autocatalysis exerts upon a system. (a) Original system configuration with numerous equiponderant interactions. (b) Same system after autocatalysis has pruned some interactions, strengthened others, and increased the overall level of system activity (indicated by the thickening of the arrows.) Corresponding matrices of topological connections indicated to the right.

hypothetical, inchoate 4- component network before autocatalysis has developed, and the lower one, the same system after autocatalysis has matured. The magnitudes of the flows are represented by the thicknesses of the arrows. To the right appear the matrices that correspond to the pattern of flows. One recognizes that the transition between matrices resembles that between Tables 7.1 and 7.2 that was presented earlier in connection with Popper's propensities.

One begins the analysis by defining the transfer of material or energy from prey (or donor) i to predator (or receptor) j as T_{ij} , where i and j range over all members of a system with n elements. The total activity of the system can be measured simply as the sum of all system processes, $T_{..} = \sum_{i,j} T_{ij}$, or what is called the "total system throughput". (Henceforth a dot in the place of any subscript will denote summation over that index.) The first aspect of autocatalysis can thus be represented as any increase in the total system throughput, much as economic growth is reckoned by any increase in Gross Domestic Product.

It is the second aspect that bears upon constraint, because the "pruning" referred to can be regarded as the appearance of additional constraints that channel flow ever more narrowly along efficient pathways—"efficient" here meaning those pathways that most effectively participate in the autocatalytic process. Another way of looking at pruning is to consider that constraints cause certain flow events to occur more frequently than others. The quantification of constraint begins by estimating the joint probability that a quantum of medium is *constrained* both to leave i and enter j as the quotient $T_{ij}/T_{..}$. It should be noted that the *unconstrained* probability that a quantum has left i can be acquired from the joint probability merely by summing the joint probability over all possible destinations. The estimator of this unconstrained probability thus becomes $T_{i.}/T_{..}$. Similarly, the unconstrained probability that a quantum enters j becomes $T_{.j}/T_{..}$. Finally, one notes how the probability that the quantum could make its way by pure chance from i to j , *without* the action of any *constraint*, would be equal to the product of the latter two frequencies, or $T_{i.}T_{.j}/T_{..}^2$.

This last probability obviously is not equal to the constrained joint probability, $T_{ij}/T_{..}$. Recalling that Tribus and McIrvine [33] defined information as "anything that causes a change in probability assignment", one may conclude that Tribus essentially equated information to constraint. Information theory, therefore, must contain clues as to how to quantify constraint. It does not, however, address information (constraint) directly. Rather it uses as

its starting point a measure of the rareness of an event, first defined by Boltzmann [34] as $(-k \log p)$, where p is the probability ($0 \leq p \leq 1$) of the given event happening and k is a scalar constant that imparts dimensions to the measure. One notices that for rare events ($p \approx 0$), this measure is very large and for very common events ($p \approx 1$), it is diminishingly small.

Because constraint usually acts to make things happen more frequently in a particular way, one expects that, on average, an unconstrained probability would be rarer than a corresponding constrained event. The more rare (unconstrained) circumstance that a quantum leaves i and accidentally makes its way to j can be quantified by applying the Boltzmann [34] formula to the probability just defined, i.e., $-k \log (T_{.j}T_{i.}/T_{..}^2)$, and the correspondingly less rare condition that the quantum is constrained both to leave i and enter j becomes $-k \log (T_{ij}/T_{..})$. Subtracting the latter from the former and combining the logarithms yields a measure of the hidden constraints that channel the flow from i to j as $k \log (T_{ij}T_{..}/T_{.j}T_{i.})$. (It is noted in passing that this quantity also measures the *propensity* for flow from i to j (Ulanowicz [35]).)

Finally, to estimate the average constraint at work in the system as a whole, one weights each individual propensity by the joint probability of constrained flow from i to j and sums over all combinations of i and j . That is,

$$AMC = k \sum_{i,j} \left(\frac{T_{ij}}{T_{..}} \right) \log \left(\frac{T_{ij}T_{..}}{T_{.j}T_{i.}} \right) \quad (6.1)$$

where AMC is the “average mutual constraint” (known in information theory as the average mutual information.)

To illustrate how an increase in AMC actually tracks the “pruning” process, the reader is referred to the three hypothetical configurations in Figure 7.7. In configuration (a) where medium from any one compartment will next flow is maximally indeterminate. AMC is identically zero. The possibilities in network (b) are somewhat more constrained. Flow exiting any compartment can proceed to only two other compartments, and the AMC rises accordingly. Finally, flow in schema (c) is maximally constrained, and the AMC assumes its maximal value for a network of dimension 4. Zorach and Ulanowicz [36] have shown how the geometric mean number of roles (trophic levels) in a flow network can be estimated as b^{AMC} , where b is the base used to calculate the logarithms in the formula for AMC.

One notes in the formula for AMC that the scalar constant, k , has been retained. Although autocatalysis is a unitary process, separate measures

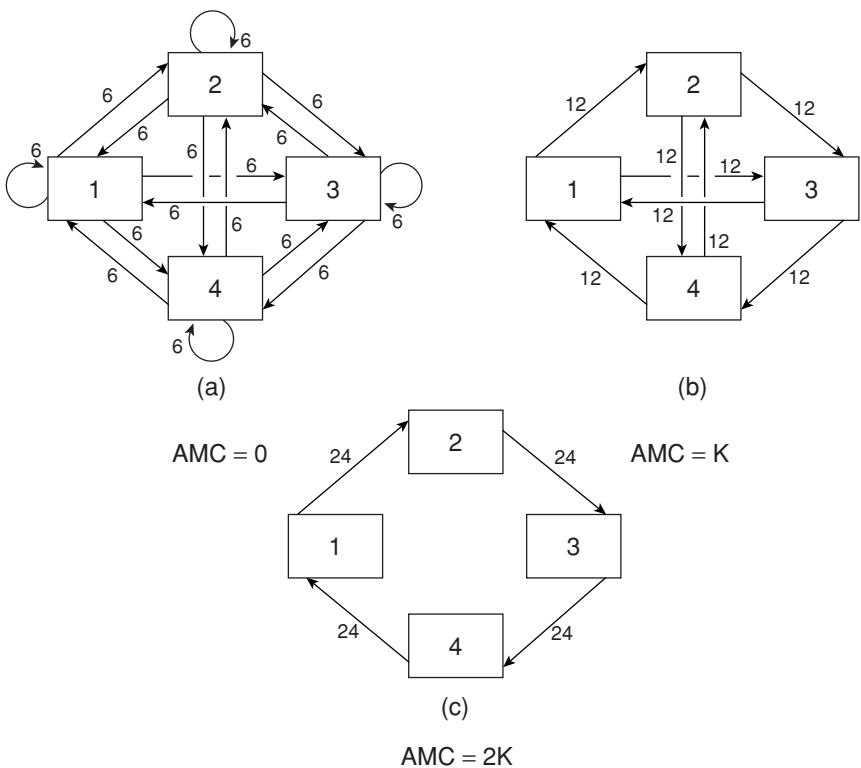


Figure 7.7. (a) The most equivocal distribution of 96 units of transfer among four system components. (b) A more constrained distribution of the same total flow. (c) The maximally constrained pattern of 96 units of transfer involving all four components.

have been defined for its two attributes. One can easily rectify this disparity and combine the measures of both attributes simply by making the scalar constant k represent the level of system activity, $T_{..}$, that is k is set equal to $T_{..}$, and the resulting product is called the system *ascendency*, A , where

$$A = \sum_{i,j} T_{ij} \log \left(\frac{T_{ij} T_{..}}{T_{.j} T_{i.}} \right). \tag{6.2}$$

In his seminal paper, “The strategy of ecosystem development”, Eugene Odum [37] identified 24 attributes that characterize more mature ecosystems. These can be grouped into categories labeled species richness, dietary specificity, recycling and containment. All other things being equal, a rise

in any of these four attributes also serves to augment the system ascendancy (Ulanowicz [35]). It follows as a phenomenological principle that “*in the absence of major perturbations, ecosystems have a propensity to increase in ascendancy.*”

It should be emphasized in the strongest terms possible that increasing ascendancy is only half of the dynamical story. Ascendancy accounts for how efficiently and coherently the system processes medium. Using the same mathematics, one can compute as well an index called the system overhead, Φ , which is complementary to the ascendancy (Ulanowicz and Norden [38]):

$$\Phi = - \sum_{i,j} T_{ij} \log \left(\frac{T_{ij}^2}{T_{.j} T_{i.}} \right). \quad (6.3)$$

Overhead quantifies the inefficiencies and incoherencies present in the system. As with the ascendancy, Zorach and Ulanowicz [36] have demonstrated how the logarithm of the geometric mean of the network link-density, LD, is equal to one-half of the overhead. That is, $LD = b^{(\Phi/2)}$. (Link-density is the effective number of arcs entering or leaving a typical node. It is one measure of the connectivity of the network.) Although the inefficiencies contributing to overhead may encumber the system’s overall performance at processing medium, they become absolutely essential to system survival whenever the system incurs a novel perturbation. At such time, the overhead comes to represent the degrees of freedom available to the system and the repertoire of potential tactics from which the system can draw to adapt to the new circumstances. Without sufficient overhead, a system is unable create an effective response to the arbitrary rigors of its environment. The configurations observed in nature, therefore, appear to be the results of a dynamical tension between two antagonistic tendencies (ascendancy vs. overhead.)

Experience has revealed that the effective numbers of roles and the connectivities of real ecosystems are not arbitrary [39]. It has long been known, for example, that the number of trophic roles (levels) in ecosystems is generally fewer than 5 (Pimm and Lawton [40]). Similarly, the effective link-density of ecosystems (and a host of other natural systems) almost never exceeds 3 (Pimm [41], Wagensberg *et al.* [42] 1990.) Regarding this last stricture, Ulanowicz [43] (2002) suggested how the May-Wigner stability criterion [44] could be re-interpreted in information-theoretic

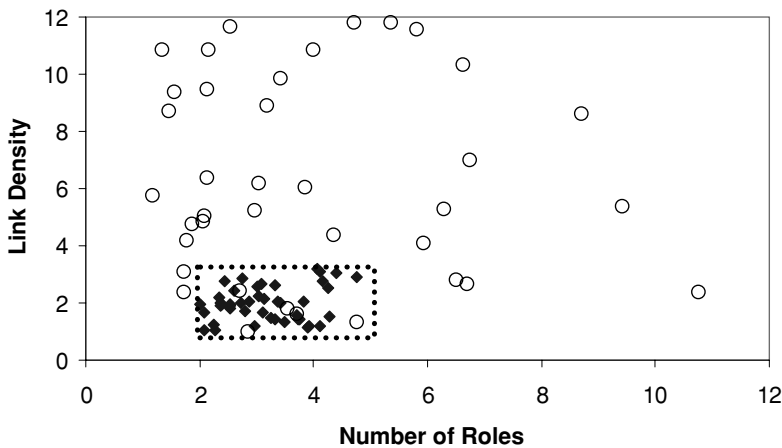


Figure 7.8. Combinations of link densities and numbers of roles pertaining to random networks (open circles) and actual ecosystems networks (solid squares.) The “window of vitality” is indicated by the dotted lines.

terms to identify a threshold of stability at $e^{e/3}$, or ca 3.015 links per node.

Limits or complexity appear quite visibly when one plots the number of roles versus the effective connectivity of a collection of 44 estimated ecosystem flow networks (Figure 7.8.) Whereas the pairs of numbers generated by randomly-constructed networks are scattered broadly over the positive quadrant, those associated with actual ecosystem networks are confined to a small rectangle near the origin. Ulanowicz [45] has labeled this rectangle the “window of vitality”, because it appears that the entire drama of ecosystem dynamics plays out within this small theatre: As mentioned above, the *endogenous* tendency of ecosystems is to drift towards the right within this window (i.e., toward ever-increasing ascendancy, or higher system performance.) At any time, however, singular events can appear as exogenous perturbations that shift the network abruptly to the left. (Whether the link- density rises or falls during this transition depends upon the nature and severity of the disturbance.) In particular, whenever the system approaches one of the outer edges of the window, the probability increases that it will fall back towards the interior. Near the top, horizontal barrier (LD = ~ 3.015 links per node) the system lacks sufficient cohesiveness and disintegrates spontaneously. As the system approaches the right-hand

frame (# roles = 4.5 to 5.0), it presumably undergoes something like a “self-organizing catastrophe” (Bak [46]) as described by Holling [47] (see also Ulanowicz [29]).

As one follows the historical trajectory of an ecosystem within the window of vitality, it is important to hold firmly in mind that any description of a history solely in terms of mechanisms and the actions of individual organisms will perforce remain inadequate. Rather, the prevailing agencies at work are the tendency of configurations of processes (sub graphs) to increase in ascendancy acting in opposition to the entropic tendency generated by complex, singular events.

7. New Constraints to Help Focus a New Perspective

It is worthwhile at this juncture to recapitulate what has been accomplished: First, the focus in ecosystem dynamics has been shifted away from the normal (symmetrical) field equations of physics and directed instead towards the origins of asymmetry in any system—the boundary constraints. It was then noted how biotic entities often serve as the origins of such constraint upon other biota, so that the kernel of ecodynamics is revealed to be the mutual (self-entailing) constraints that occur within the ecosystem itself. Then a palpable and measurable entity (the network of material/energy exchanges) was identified upon which the myriad of (mostly hidden) constraints could write its signature. Finally, a calculus was developed that could quantify the effects of all the hidden constraints. As a result, by following changes in the ascendancy and overhead of an ecosystem flow network, one is focussing squarely upon that which makes ecodynamics fundamentally different from classical dynamics.

By many accounts, the Enlightenment started in earnest with Newton’s publication of *Principia*, which provided a quantitative basis for classical dynamics. In the years that followed, numerous thinkers built around Newtonian dynamics a supporting metaphysic that for the last three centuries has strongly guided how one is to look at nature. It is only fair to ask how well does that metaphysic support the emerging ecodynamics that have just been described (Ulanowicz [48])? To provide a basis for comparison, one must first describe the Newtonian metaphysic as it appeared at its zenith.

Depew and Weber [49] have identified four postulates under which Newtonian investigations were pursued during the early 19th Century:

- Newtonian systems are causally *closed*. Only mechanical or material causes are legitimate.
- Newtonian systems are *deterministic*. Given precise initial conditions, the future (and past) states of a system can be specified with arbitrary precision.
- Newtonian systems are *reversible*. Laws governing behavior work the same in both temporal directions.
- Newtonian systems are *atomistic*. They are strongly decomposable into stable least units, which can be built up and taken apart again.

To Depew and Weber's list may be added a fifth article of faith (Prigogine and Stengers [50], Ulanowicz [29]), namely that

- Physical laws are *universal*. They apply everywhere, at all times and over all scales.

Early in the 19th Century, the notion of reversibility had already been challenged by Sadi Carnot's thermodynamical elaboration of irreversibility and several decades later by Darwin's historical narrative. The development of relativity and quantum theories early in the 20th Century worked to subvert even further the assumptions of universality and determinism, respectively. Despite these problems, many in biology (and especially in ecology) continue to operate under the mechanistic umbrella just delimited.

Given the ground that has been covered here, it becomes apparent that the Newtonian metaphysic accords rather poorly with ecodynamics. In fact, the new dynamics appear to be dissonant with *each* of the five Newtonian precepts. To wit:

1. Ecosystems are not causally closed in that they appear to be *open* to the influence of non-mechanical agency. Spontaneous events may occur at any level of the hierarchy at any time. Efficient (or mechanical) causes usually originate at scales inferior to that of observation, and their effects propagate upwards. Formal agencies appear at the focal level; and final causes exist at higher levels and propagate downwards (Salthe [6]; Ulanowicz [29])
2. Ecosystems are not deterministic machines. They are *contingent* in nature, and such contingency is often singular. Biotic actions resemble propensities more than mechanical forces.

3. The realm of ecology is *granular*, rather than universal. Models of events at any one scale can explain matters at another scale only in inverse proportion to the remoteness between them. On the other hand, the domain within which irregularities and perturbations can damage a system is usually circumscribed. Chance does not necessarily unravel a system.
4. Ecosystems, like other biotic systems, are not reversible, but *historical*. Irregularities often take the form of (often singular) discontinuities, which degrade predictability into the future and obscure hindcasting. The effects of past discontinuities are often retained (as memories) in the material and kinetic forms that result from adaptation. Time takes a preferred direction or telos in ecosystems—that of increasing ascendancy.
5. Ecosystems are not easily decomposed; they are *organic* in composition and behavior. Propensities never exist in isolation from other propensities, and communication between them fosters clusters of mutually reinforcing propensities to grow successively more interdependent. Hence, the observation of any component in isolation (if possible) reveals regressively less about how it behaves within the ensemble.

Although this ecological worldview may at first blush seem wholly revolutionary, it actually follows Popper's *evolutionary* leads and thereby retains some connections with the orthodox and the classical. For example, because propensities are generalizations of Newtonian forces, it can be shown how the principle of increasing ascendancy resembles a generalization of Newtonian law upwards into the macroscopic realm, in a way similar to how Schroedinger's wave equation is an extension of Newton's second law downwards into the netherworld of quantum phenomena (Ulanowicz [48])

Such continuity notwithstanding, it would be a major distortion to claim that the ecological metaphysic describes a new mechanics (in much the same manner that quantum physics is often mistakenly still referred to as "quantum mechanics", despite the fact that there is virtually nothing mechanical about the phenomena.) As Casti has admonished, it is past time to make a clean break with the vision of "*natura cum machina*" (Dennett [51]). If it doesn't look like a machine, if it doesn't act like a machine, if it doesn't smell like a machine, why then persist in calling it a machine? Such procrustean nostalgia only fosters a highly distorted vision of the natural world. Metaphors and methods are emerging that are far more effective and appropriate to charting the pathways that the living world has blazed

for itself (Ulanowicz [52]) The era of climbing trees is passing; the time is well-nigh to move beyond rocket science.

Acknowledgements

The author was supported in part by the National Science Foundation's Program on Biocomplexity (Contract No. DEB-9981328.) Mr. Michael Zickel assisted with some of the figures. The author would like to thank Drs. Lemont Kier and Danail Bonchev for their encouragement to prepare this essay.

References

1. J. E. Cohen, *Science* **172**, 674-675 (1971).
2. R. Strand, *Emergence* **4**, 164-183 (2002).
3. S. Goerner, personal communication.
4. J. Casti, Newton, Aristotle, and the Modeling of Living Systems, in *Newton to Aristotle*, J. Casti and A. Karlqvist (eds.), Birkhaeuser, NY (1989) pp. 47-89.
5. J. Kay and E. D. Schneider, *Alternatives* **20**, 32-38 (1994).
6. S. N. Salthe, *Evolving Hierarchical Systems: Their Structure and Representation*. Columbia University Press, New York (1985).
7. R. Rosen, *Life Itself: A Comprehensive Inquiry into the Nature, Origin and Foundation of Life*, Columbia University Press, NY (1991).
8. D. Mickulecky, *Systems Research and Behavioral Science* **17**, 419- 432 (2000).
9. J. L. Casti, *Why the Future Happens, The Second International Biennial Seminar on the Philosophical, Methodological and Epistemological Implications of Complexity Theory*, International Convention Center, La Habana, Cuba, Jan 7-11, (2004).
10. E. M. Minarik, *Little Bear*, Harper and Row, New York (1957).
11. D. Brooks, *Evolution in the Information age: Three Converging Viewpoints. The Second International Biennial Seminar on the Philosophical, Methodological and Epistemological Implications of Complexity Theory*, International Convention Center, La Habana, Cuba, Jan 7-11 (2004).
12. G. Stent, Strength and Weakness of Genetic Approach to the Development of the Nervous System. *Ann. Rev. Neurosci.* **4**, 16-194 (1981).
13. R. Lewin, *Science* **224**, 1327-1329 (1984).
14. R. C. Strohman, *California Monthly* **111**, 24-27 (2001).
15. A. Noether, *Gesammelte Abhandlungen*, Nathan Jacobson (ed.), Springer Verlag, New York (1983).
16. C. F. Stevens, *The Six Core Theories of Modern Physics*, MIT Press, Cambridge, MA 1995.
17. S. E. Joergensen and H Mejer, *Ecol. Model.* **7**, 169-189 (1979).
18. J. Engleberg and L. L. Boyarsky, *Am. Nat.* **114**, 317-324 (1979).
19. R. E. Ulanowicz, *Growth and Development: Ecosystems Phenomenology*, Springer-Verlag, New York (1986).

20. R. E. Ulanowicz, *Order and Fluctuations in Ecosystem Dynamics, The Second International Biennial Seminar on the Philosophical, Methodological and Epistemological Implications of Complexity Theory*, International Convention Center, La Habana, Cuba Jan, 7-11 (2004).
21. W. M. Elsasser, *American Scientist* **57**, 502-516 (1969).
22. K. R. Popper, *A World of Propensities*, Thoemmes, Bristol (1990).
23. M. Eigen, *Naturwiss* **58**, 465-523 (1971).
24. H. Haken, *Information and Self-Organization*. Springer-Verlag, Berlin, (1988).
25. H. R. Maturana and F. J. Varela, *Autopoiesis and Cognition: The Realization of the Living*, D. Reidel, Dordrecht (1980).
26. S. Kauffman, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*, Oxford University Press, New York (1995).
27. D. L. DeAngelis, W. M. Post, and C. C. Travis, *Positive Feedback in Natural Systems*, Springer-Verlag, New York (1986).
28. R. E. Ulanowicz, *Ecological Modelling* **79**, 49-57 (1995).
29. R. E. Ulanowicz, *Ecology, the Ascendent Perspective*. Columbia University Press, New York (1997).
30. R. Rosen, Information and Complexity, in *Ecosystem Theory for Biological Oceanography*, RE Ulanowicz, T Platt (eds.), *Canadian Bulletin of Fisheries and Aquatic Sciences* **213**, 221-233 (1985).
31. T. F. H. Allen and T. B. Starr, *Hierarchy*, University of Chicago Press, Chicago (1982).
32. S. N. Salthe, *Development and Evolution: Complexity and Change in Biology*, MIT Press, Cambridge (1993).
33. M. Tribus and E. C. McIrvine, *Sci. Am.* **225**, 179-188 (1971).
34. L. Boltzmann, *Wien. Ber.* **66**, 275-370 (1872).
35. R. E. Ulanowicz, *Growth and Development: Ecosystems Phenomenology*. Springer-Verlag, New York (1986).
36. A. C. Zorach and R. E. Ulanowicz, *Complexity* **8**, 68-76 (2003).
37. E. P. Odum, *Science* **164**, 262-270 (1969).
38. R. E. Ulanowicz and J. S. Norden, *International Journal of Systems Science* **1**, 429-437 (1990).
39. R. E. Ulanowicz, Ecological Network Analysis: An Escape from the Machine, in *Aquatic Food Webs*, A. Belgrano, U. M. Scharler, J. Dunne, and R. E. Ulanowicz (eds.), Oxford University Press, Oxford, UK, (2005), pp. 201-207.
40. S. L. Pimm and J. H. Lawton, *Nature* **268**, 329-331 (1977).
41. S. L. Pimm, *Foodwebs*, Chapman and Hall, London (1982).
42. J. Wagensberg, A. Garcia, and R. V. Sole, *Bull. Math. Biol.* **52**, 733-740 (1990).
43. R. E. Ulanowicz, *BioSystems* **64**, 13-22 (2002).
44. R. M. May, *Nature* **238**, 413-414 (1972).
45. R. E. Ulanowicz, Limitations on the Connectivity of Ecosystem Flow Networks, in *Biological Models*, A. Rinaldo and A. Marani (eds.), Istituto Veneto de Scienze, Lettere ed Arti, Venice (1997) pp.125-143.

46. P. Bak, *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, New York (1996).
47. C. S. Holling, The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change, in *Sustainable Development of the Biosphere*, W.C. Clark, and R.E. Munn (eds.), Cambridge University Press, Cambridge, UK (1986) p. 292-317.
48. R. E. Ulanowicz, *BioSystems* **50**, 127-142 (1999).
49. D. J. Depew and B. H. Weber, *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection*, MIT Press, Cambridge, MA (1994).
50. I. Prigogine and I. Stengers, *Order out of Chaos: Man's New Dialogue with Nature*, Bantam, New York (1984).
51. D. C. Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, Simon and Schuster, New York (1995).
52. R. E. Ulanowicz, *Ludus Vitalis* **9**, 183-204 (2001).

AUTHOR INDEX

- Abraham, R., 117, 118
Adam, G., 284
Ajayan, P. M., 38
Albert, R., 170, 181, 221, 229, 253
Alberts, B., 164
Aldana, M., 186, 187
Aldana-González, M., 182
Allen, M. P., 246, 247
Allen, T. F. H., 251, 317
Altenberg, L., 166
Ancel, L. W., 166
Anderson, P. W., 239
Andriotis, A. N., 43
Arbib, M., 118
Artavanis-Tsakonas, S., 67
Asai, R., 71
Auslander, D. M., 115
Ayers, F., Hr., 5
Aylsworth, A. S., 76
- Babic, D., 5
Bak, P., 324
Balaban, A. T., 4, 7, 37, 41, 43
Barabási, A. L., 170, 221, 226, 253
Barone, R., 1
Barwise, J., 102
Bar-Yam, Y., 251
- Bateson, W., 65, 90
Beckenbach, E. F., 29
Belew, R. K., 250
Bellairs, R., 67
Bellman, R., 29
Ben-Jacob, E., 89
Bertalanffy, L. V., 89
Bertz, S. H., 1, 27, 192, 207, 208
Beveridge, D. L., 239
Bhalla, U. S., 256
Biro, L. P., 43
Blackwell, W. A., 115
Boissonade, J., 79
Bollobás, B., 158
Boltzmann, L., 320
Bonchev, D., 1, 2, 11, 191, 192, 198, 203, 204, 205, 207, 208, 209, 210, 211, 219, 225, 229, 230, 253, 290
Borisuk, M. T., 53
Boyarsky, L. L., 307
Branford, W. W., 76
Branin, H., Jr., 118, 126
Breedveldt, P. C., 115
Broadie, K., 184
Brooks, D., 304
Brown, S. J., 80
Bytautas, L., 27

- Cable, M. B., 116
 Calenbuhr, V., 239
 Callen, B., 121, 135
 Campos-Ortega, J. A., 67
 Caplan, S. R., 121, 136, 143
 Castets, V., 76
 Casti, J. L., 303
 Chanon, M., 1
 Chen, Y., 76
 Cheng, C. K., 274, 275, 281, 291
 Christen, B., 70
 Chua, L. O., 114, 123, 129, 130, 133
 Clements, D., 74
 Cluzel, P., 186
 Clyde, D. E., 80
 Cohen, J. E., 303
 Cohen, S. M., 70
 Combs, A., 239
 Comper, W. D., 89
 Cooke, J., 66, 90
 Cooke, K. L., 240, 252
 Cornish-Bowden, A., 85
 Cotton, F. A., 36, 37
 Crespi, V. H., 43
 Cruziat, P., 116
 Curran, P. F., 98, 121
 Cvetković, D. M., 5

 Davidson, E. H., 253
 Dawes, R., 80
 DeAngelis, D. L., 311
 Decker, P., 255
 Defay, R., 121
 Degn, H., 255
 deGroot, S. R., 121, 143
 Delbruck, M., 284
 Dennett, D. C., 326
 Depew, D. J., 325
 DeRusso, P. M., 129
 DeSimone, A., 143
 Desoer, C. A., 115
 Devillers, J., 4
 Diudea, M. V., 43
 Dorogovtsev, S. N., 170, 221, 226

 Dress, W., 98
 Dresselhaus, M. S., 43
 Dubrulle, J., 65, 66
 Dunlap, B. I., 43
 Dunne, J. A., 226, 228, 230

 Ebbesen, T. W., 38
 Eckmann, J.-P., 253
 Eigen, M., 285, 311
 El-Basil, S., 27
 Elowitz, M. B., 53
 Elsasser, W. M., 308
 Engleberg, J., 307
 Entchev, E. V., 70
 Epstein, I. R., 79
 Espadaler, J., 165
 Essig, A., 121, 136

 Featherstone, D. E., 184
 Feher, J. J., 116
 Fell, D. A., 221
 Ferell, J. E., 290, 296
 Fidelman, M. L., 116, 140
 Figueras, J., 4
 Fitts, 121
 Fontana, W., 166
 Forgacs, G., 89
 Frasch, M., 79
 Friedman, N., 221
 Frisch, H. L., 71

 Gavin, A. C., 221, 222, 223, 224, 226
 Ge, M. H., 37, 43
 Gebben, V. D., 115
 Gehring, W. J., 83
 Gell-Mann, M., 202
 Gerhart, J., 74
 Gianscotti, F. G., 59
 Gilbert, S. F., 49, 51, 65, 70
 Giot, L., 221
 Goerner, S., 303
 Goldstein, L. J., 116
 Goodwin, B. C., 54, 89
 Gordeeva, K., 192

- Gordon, M., 208
 Green, J., 70
 Gurdon, J. B., 59, 64, 90
 Gutfreund, H., 239
 Gutin, A. M., 165
 Gutman, I., 1, 5, 6, 211
- Haile, J. M., 246, 247
 Haken, H., 311
 Hamada, H., 76
 Harary, F., 193, 198
 Harding, K., 83
 Harland, R., 74
 Harrison, L. J., 84
 Hatsopoulos, G. N., 121
 Hentschel, H. G. E., 71
 Herndon, W. C., 208
 Hess, B., 255
 Ho, R. K., 67
 Hoffman, D. D., 102
 Hoffman, R., 3
 Holley, S. A., 67
 Holling, C. S., 324
 Hollon, T., 53
 Horgan, J., 101, 103, 114, 138
 Horno, J. C., 116
 Howard, K., 79
 Howell, J. R., 116
 Huang, C. Y. F., 290, 296
 Huf, E. G., 116
- Ihara, S., 43
 Ijima, S., 38
 Ingham, P., 79
 Irvine, K. D., 79
 Ish-Horowicz, D., 83
 Ito, T., 171, 173, 221
 Itow, T., 77
 Iyengar, R., 256
- Jablonka, E., 54
 Jaszczak, J. A., 43
 Jeong, H., 221
 Jiang, T., 71
- Joergensen, S. E., 306
 John, P. E., 1
 Jorgensen, W. L., 239, 246, 247
 Juan, H., 76
- Kamenska, V., 192
 Kaneko, K., 60, 62, 64, 67, 91
 Karabunarliev, S., 225
 Karamata, J., 28
 Karnopp, D., 115, 118
 Karr, T. L., 79
 Kastler, H., 191
 Katchalsky, A., 98, 121, 132, 141
 Kauffman, S., 311
 Kauffman, S. A., 54, 181, 182
 Kay, J., 303
 Kay, J. J., 98
 Kedem, O., 132, 136, 141
 Keenan, J. H., 121
 Keller, A. D., 55, 56, 57, 67, 85, 90, 91
 Kerckel, S. W., 147
 Kerszberg, M., 70
 Kier, L. B., 256, 274, 275, 277, 278, 280, 281, 284, 289, 290, 291
 Kilian, J., 43
 Kirby, E. C., 43
 Kirchhoff, R., 127
 Klein, D. J., 5, 27, 37, 38
 Koch, C., 221
 Koenig, H. E., 115
 Kolmogorov, A. N., 191
 Kondo, S., 71
 Kostoff, R. N., 38
 Krätschmer, W., 38
 Kroto, H. W., 38
 Kurusawa, C., 60, 62
- Laidboeur, T., 20, 24, 34
 Lam, Y., 130
 Lamb, M. J., 54
 Lammert, P. E., 43
 Lander, A. D., 70
 Laurent, G., 221
 Lawrence, P. A., 77, 79

- Lawton, J. H., 322
 Leach, A. R., 246
 Lee, T. I., 221
 Leibler, S., 53
 Lengyel, I., 79
 Lercher, M. J., 53
 Levin, S. A., 239
 Levine, H., 89
 Levine, M., 79
 Lewin, R., 305
 Lewis, J., 67, 69, 90
 Li, S., 171
 Lin, P., 129, 133
 Linshitz, H., 191
 Liu, J., 43
 Lovasz, L., 5
 Lukovits, I., 43

 Ma, H., 221
 MacDonald, N., 252
 MacFarlane, G. J., 115
 Madan, R. N., 114, 130
 Manes, E. G., 118
 Mangold, H., 71
 Mannervik, M., 55
 Margolus, N., 257
 Marsden, J. E., 118
 Maslov, M., 172
 Maslov, S., 172, 229
 Mason, S. J., 132
 Maturana, H. R., 112, 146, 311
 May, J. M., 116
 May, R. M., 322
 Mazur, P., 121, 143
 McDowell, N., 70
 McIrvine, E. C., 319
 Meinhardt, H., 71, 72, 74, 75, 84, 89, 90
 Meir, E., 90
 Mejer, H., 306
 Mendes, J. F. F., 170, 221
 Merz, K. M., Jr., 37
 Mierson, S., 116
 Mikhailov, A. S., 239

 Mikulecky, D. C., 99, 101, 103, 114, 115, 116, 119, 122, 130, 133, 134, 136, 138, 140, 146, 147, 153, 303
 Miller, D. G., 137
 Milo, R., 253
 Minarik, E. M., 303
 Minoli, D., 191
 Mintz, E., 116
 Mitchell, M., 250
 Monk, N. A., 67, 90
 Montoya, J. M., 253
 Morgan, H. L., 4, 17, 197
 Morisco, C., 59
 Moss, L., 102
 Mowshowitz, A., 2, 191
 Muirhead, R. F., 25
 Müller, G. B., 54, 77, 89, 91
 Muratov, C. B., 79

National Science Foundation, 245
 Neixner, J., 115
 Neuman, M. E. J., 199, 200
 Newman, M. E. J., 159, 172
 Newman, S. A., 54, 71, 76, 77, 83, 89, 91
 Nicolić, N., 205, 211
 Nicolis, G., 255
 Nieuwkoop, P. D., 70
 Nijhout, H. F., 71
 Nikolic, S., 1, 7, 15, 192
 Noether, A., 305
 Nonaka, S., 76
 Norden, J. S., 322
 Noyes, R. M., 255
 Nusslein-Volhard, C., 79

 Oates, A. C., 67
 Odum, E. P., 321
 Oken, D. E., 116
 Olsen, L. F., 255
 Onsager, L., 131, 136
 Oster, G. F., 98, 115, 118, 120, 122, 126
 Othmer, H. G., 181
 Ouyang, Q., 76

- Palmeirim, I., 65, 66, 90
 Patel, N. H., 77, 80
 Pelikan, J., 5
 Penfield, P., Jr., 120
 Perelson, S., 115, 118
 Peusner, L., 98, 115, 116, 119, 136
 Pimm, S. L., 322
 Platt, J. R., 207
 Playšić, D., 1, 20, 30, 205
 Plessner, T., 255
 Polansky, O. E., 2, 11, 192
 Popper, K., 98
 Popper, K. R., 309
 Pourquié, O., 65, 66, 84
 Prideaux, J., 116
 Prigogine, I., 121, 255, 325
 Primas, H., 1, 3
 Primmatt, D. R., 67

 Randić, M., 1, 3, 4, 5, 6, 7, 8, 20, 27, 28, 30, 205
 Rashevsky, N., 89, 103, 191, 252
 Ravasz, E., 166, 253
 Razinger, M., 4, 17
 Reinitz, J., 85, 91
 Renfrew, C., 240, 252
 Rideout, V. C., 115
 Roe, P. H., 115
 Rosen, R., 98, 99, 101, 103, 121, 126, 138, 143, 252, 303, 306, 317
 Rosenberg, R. C., 115, 118
 Rouvray, D. H., 205
 Ruch, E., 5, 6
 Rücker, C., 1, 4, 192, 209, 211, 212
 Rücker, G., 1, 4, 192, 209, 211, 212
 Ruoslahti, E., 59
 Rypins, E. B., 116

 Sakuma, R., 76
 Salazar-Cuidad, I., 78, 81, 82, 83, 84, 85, 86, 87, 88, 91
 Salthe, S. N., 251, 303, 317, 325
 Satorras-Pastor, R., 253
 Sattler, K., 37, 43

 Sauer, F. A., 115
 Savageau, M. A., 239
 Scantlebury, G. R., 208
 Schier, A. F., 76, 83
 Schmalhausen, I. I., 83
 Schneider, E. D., 98, 303
 Schroeder, M., 257
 Schulte-Merker, S., 74
 Scuseria, G. E., 43
 Segel, L. A., 239
 Seither, R. L., 116, 134
 Seitz, W. A., 192, 208, 211, 230
 Selkov, E. E., 255
 Shannon, C., 2, 191
 Sharma, K. R., 255
 Shaw, C. D., 117
 Slack, J., 70
 Small, S., 80, 81
 Smith, J. C., 74
 Sneppen, K., 172, 229
 Solé, R. V., 155, 166, 175, 177, 181, 253
 Solnica-Krezel, L., 72, 76
 Sommer, T. J., 192, 230
 Spemann, H., 71
 St. Johnston, D., 79
 Stadler, N. M., 166
 Standley, H. J., 59
 Starr, T. B., 251, 317
 Stengers, I., 325
 Stent, G., 304
 Stern, C. D., 67
 Stevens, C. F., 305
 Stollewerk, A., 79
 Strand, R., 303
 Strogatz, S. H., 52, 53, 54, 163, 213
 Strohmaier, R. C., 305
 Sulis, W., 239
 Sun, B., 73
 Swinney, H., 76

 Takke, C., 67
 Talley, D. B., 116
 Teleman, A. A., 70
 Tellegen, D. H., 120

- Terrones, M., 43
 Testa, B., 256
 Thakker, K. M., 116, 134
 Thellier, M., 116
 Thoma, J. U., 115
 Thomas, R., 116
 Thomas, S. R., 128, 133
 Tildesley, D. J., 246
 Tirado-Rives, J., 247
 Tisza, L., 121
 Toffoli, T., 257
 Tong, A. H. Y., 185, 221
 Townend, M. S., 239
 Treacy, M. M. J., 43
 Tribus, M., 319
 Trinajstić, N., 192, 193, 203, 209, 210, 211
 Trucco, E., 191
 Truesdell, 121
 Tuinenga, P. W., 128, 133
 Turing, A., 72, 73, 90
 Tyson, J. J., 53

 Ulam, S. M., 257
 Ulanowicz, R. E., 307, 308, 313, 316, 320, 322, 323, 324, 325, 326, 327

 Van Obberghen-Schilling, E., 73
 Varela, F. J., 112, 146, 311
 Vázquez, A., 176, 178
 Vichniac, G. Y., 257
 Vogt, K., 27
 von Dassow, G., 90
 von Neumann, J., 257

 Waddington, C. H., 83, 286
 Wagensberg, J., 322

 Wagner, A., 221
 Wagner, G. P., 166
 Waldrop, M. M., 254
 Walz, D., 116
 Watts, D., 163
 Watts, D. J., 213
 Weaver, W., 2, 191
 Weber, B. H., 325
 Welch, G. R., 284, 285
 Wells, H. G., 256
 Welz, G., 27
 Weng, G., 221
 White, J. C., 116, 134
 Wiener, H., 1, 211
 Wieschaus, E., 79
 Winfree, A. T., 89
 Witten, T. M., 239, 243, 245, 246, 252, 253
 Wolfram, S., 257
 Wolpert, L., 70
 Wright, W. F., 192, 230
 Wyatt, J. L., 115, 128, 134

 Yamamoto, M., 76
 Yomo, T., 60, 62, 64, 91
 Yook, S. H., 226
 Yost, H. J., 76

 Zamfirescu, C. M., 27
 Zeeman, E. C., 66, 90
 Zengl, A. P., 221
 Zhang, G., 43
 Zhu, H., 37
 Zhu, H. Y., 43
 Zimmermann, H. J., 132
 Zorach, A. C., 320, 322
 Zuse, K., 257

SUBJECT INDEX

- A/D index, 213–215, 217
- Absolute gravity, 271
- Accessible connectedness, 219–220
- Activators, 55
- Acyclic complexity, 5
- Adjacency matrix, 195–196
 - leading eigenvalue of, 7–16
- Adjacent vertices, 195
- Adjusted average distance, 219
- Admittance matrix, 129
- Agents, 250–251
- Algorithmic information, 191
- Algorithms, 108
- AMC (average mutual constraint), 320–321
- Analog models of network thermodynamics, 121
- Arcs, 194
- Armchair nanotubes, 39
- Ascendency, 321
- Assortativeness of networks, 172
- Asynchronous cell movement, 264–266
- Atom environments, local, 25–29
- Atomic complexity index, 23
- Augmented valence complexity index, 16–20, 30
- Autocatalysis, 73
 - non–mechanical behavior and, 311–316
- Autopoiesis, 146
- Autopoietic unity, 112
- Autoregulatory transcription factor network model, Keller, 55–59
- Average degree, 158
- Average edge complexity, 206
- Average graph distance, 198
- Average mutual constraint (AMC), 320–321
- Average path length, 159–161
- Average vertex degree, 196
- Axis formation
 - left–right symmetry and, 71–72
 - Meinhardt’s models for, 72–76
- B index, 215–218
- Basin of attraction, 54
- Biochemical networks, modeling, 289–297
- Biochemical state of cells, 52
- Biochemical systems, complex, cellular automata models of, 237–298
- Biological development and evolution, complex chemical systems in, 49–51

- Biological networks, complexity
 - estimates of, 221–230
- Biology, relational systems theory in, 141–144
- Biosystems, complex, modeling emergence in, 257–274
- Blastulae, 51
- Bond graphs, 118
- Boolean network, random (RBN), 181–183
- Boron, 36
- Boundary conditions, zero–flux, 86
- Brain, 101
- Branching, 5
 - complexity and, 4–7
 - degree of, 6, 11
 - molecular, 5, 6–7
- Breaking rules, 269
- Canalizing selection, 83
- Capacitance, electrical, 124–125
- Capacitor, 122
- Carbon, 36
- Carbon nanotubes, *see* Nanotubes
- Causality, 99, 316–317
- Cayley–Hamilton theorem, 5
- Cell differentiation, dependence of, 59–64
- Cell division cycle, 52
- Cell movement, 262–267
- Cell movement rules, 267–273
- Cell rotation, 271, 273
- Cell shape, 259
- Cell theory, 146
- Cell type diversification, multistability in, 53–55
- Cell types, 261–262
- Cells, 155, 258–262
- Cellular automata, 138, 257–262
- Cellular automata models
 - collection of data in, 273–274
 - of complex biochemical systems, 237–298
 - examples of, 274–297
 - biochemical networks, 289–297
- Cellular automata models (*cont.*)
 - chreode theory of diffusion in water, 283–289
 - diffusion in water, 280–283
 - molecular bond interactions, 277–280
 - water structure, 275–277
- Chaotic systems, 115
- Chemical dynamics, insect segmentation and, 80–83
- Chemical oscillations, somitogenesis and, 65–66
- Chemical reaction networks, simulation of, 134
- Chemistry
 - models in, 246–248
 - relational systems theory in, 141–144
- Chreode theory of diffusion in water, 283–289
- Chreodes, 280–281
- Church–Turing thesis, 108
- Circular reasoning, 3
- Circularity of context dependence, 97–98
- Clustering, hierarchical, in contact maps, 166–169
- Clustering coefficient, 161–162, 196–197
 - second, 197
- Combinatorial complexity, 191
- Community effect, 59–60
- Comparability inequalities, Muirhead, 28
- Compartmental systems, mass transport in, 134
- Complete graphs, 194
- Complex chemical systems in biological development and evolution, 49–51
- Complex reality, 147
- Complexity, 2, 90, 144
 - acyclic, 5
 - attempts to simplify, 97–148
 - branching and, 4–7
 - of carbon nanotubes, 36–43
 - combinatorial, 191

- Complexity (*cont.*)
 of complexity concept, 3–4
 compositional, 191
 defining, 248–257
 of ecodynamics, 303–327
 general principles of, 248–257
 global, 24
 molecular, 1
 network, *see* Network complexity
 of smaller fullerenes, 20–24
 of smaller molecules, 7–16
Complexity concept, complexity of, 3–4
Complexity estimates of biological and ecological networks, 221–230
Complexity index
 atomic, 23
 augmented valence, 16–20, 30
 molecular, 23
Complexity index B, 215–218
Complexity measures, 1
 combined, 213–218
 desirable properties for, 2
Complexity theory, 114, 192
 emergence in, 139–148
Complexity vectors, 35
Components, 4, 159
 functional, 106
 graph, 194
Compositional complexity, 191
Connected graph, 194
Connectedness, 196
 accessible, 219–220
Connectivity, 4
 extended, 17, 197
 overall (OC), 210–211
Constitutive laws for physical systems, 122–123
Constraint(s)
 incorporated, measuring effects of, 306–307
 new, in ecosystems, 324–327
 quantifying, in ecosystems, 318–324
Contact graphs, protein structure and, 164–169
Contact maps, hierarchical clustering in, 166–169
Context dependence
 circularity of, 97–98
 self reference and, 102–103
Contingency, ecosystems and, 307–311
Correlation profiles, 172–175
Crude symmetry (CS), 33, 34
CS (crude symmetry), 33, 34
Curie's principle, 142–143
Cycle, 194
Cyclic graphs, 194

DDSs (distance degree sequences), 20–24, 198
Degree, 157
 of branching, 6, 11
Degree distribution, 158
Dependency networks, 253
Determination, 51
Deterministic cell movement rules, 266–267
Developmental mechanisms, evolution of, 76–89
Developmental robustness, evolution of, 83–89
Differential gene expression, 49
Differentiation, 52
Diffusion in water, 280–283
Directed graphs, 127, 194, 253
Disconnected graph, 194
Distance, 198
Distance degree distribution, 198
Distance degree sequences (DDSs), 20–24, 198
Distance in–degrees, 200
Distance magnitude distribution, 198
Distance matrix, 198
Distance out–degrees, 200
Dynamics, 125
 Newtonian, 112–114

Ecodynamics, 303
 complexity of, 303–327

- Ecological networks, complexity
 - estimates of, 221–230
- Ecosystems
 - contingency and, 307–311
 - new constraints in, 324–327
 - quantifying constraints in, 318–324
- Edges, 157, 193
- Electric circuits, network
 - thermodynamics and, 119–120
- Electrical capacitance, 124–125
- Embryo pattern formation, reaction–
 - diffusion mechanisms and, 70–76
- Embryogenesis, 49
- Emergence, 256
 - in complexity theory, 139–148
 - modeling, in complex biosystems, 257–274
- Entailment, 110
- Entropy
 - of information, 202
 - Shannon, 204
- Epigenetic inheritance, 54
- Epigenetic multistability, 55–59
- Epigenetic system, 84
- Euler's theorem, 194
- Extended connectivity, 17, 197
- External world, human mind and, 99–100

- Fick's law, 123
- Food webs, 226–230
- Formal description of networks, 126–128
- Formal system, 99
- Fragmentability, 108
- Fullerenes
 - smaller, complexity of, 20–24
 - symmetry and, 29–34
- Functional components, 106

- Gene duplication, 175
- Gene expression, differential, 49
- Gene networks, 180–187
- Genericity, 107
- Giant component, 159
- Global complexity, 24

- Global edge complexity, 206
- Gordon–Scantlebury index, 208
- Graph center, 198, 200
- Graph complexity, 2
- Graph components, 194
- Graph diameter, 198
- Graph distances, 198–200
- Graph theory, 193
 - basic notions in, 193–194
- Graphitic cones, 43
- Graphs, 157–158, 253
 - bond, 118
 - complete, 194
 - contact, protein structure and, 164–169
 - directed, 127
 - linear, 126
 - networks as, 193–201
 - random, 158
 - simple, 194
 - weighted, 201
- Gravity, 113

- Hasse diagrams, 10
- Heaviside function, 86
- Helicity, 39
 - of nanotubes, 38–43
- Hierarchical clustering in contact maps, 166–169
- Hierarchy, 251–254
- Hill function, 85
- Hopf instability, 79
- Human mind, external world and, 99–100
- Hydrophobic effect, 284–285

- In–adjacency, 196
- In–center, 200
- In–component, 200
- In–degree, 157, 196
- Incidence matrix, 127
- Incorporated constraints, measuring effects
 - of, 306–307
- Induction, mesoderm, 70

- Inductor, 122
- Inferences, 100
- Information, 104–105
 - algorithmic, 191
 - entropy of, 202
 - topological, 191
- Information content, 2
- Information indices, 192
- Insect segmentation, 77–80
 - chemical dynamics and, 80–83
- Interactions, 242–243
- Isologous diversification, 60
- Isologous diversification model,
 - Kaneko–Yomo, 59–64
- Joining parameter, 267, 268
- Kaneko–Yomo isologous diversification model, 59–64
- Keller autoregulatory transcription factor network model, 55–59
- Kirchhoff’s effort law, 128
- Kirchhoff’s flow law, 127–128
- Kirchhoff’s laws, 119
- Knowledge, quest for, 101
- Law of conservation of energy, 114
- Law of inertia, 113
- Leading eigenvalue of adjacency matrix, 7–16
- Left–right symmetry, axis formation and, 71–72
- Lewis–Monk model of somitogenesis oscillator, 66–69
- Linear graphs, 126
- Linear multiports, 130–133
- Linear resistive networks, 128–130
- Link deletion, 178
- Link–density, 322
- Links, 157, 193
- Local atom environments, 25–29
- Loop, 193, 196
- Machines, organisms versus, 108–112
- MAPK cascade signaling pathway, 290–297
- Mass transport in compartmental systems, 134
- Mathematics of science, 117–118
- Mechanisms, relational models of, 112
- Mechanistic theories, 115
- Meinhardt’s models for axis formation, 72–76
- Memristor, 123
- Mesoderm induction, 70
- Metabolism–Repair [M, R] system, 104
- Mind, human, external world and, 99–100
- Model, 99
- Modeling process, 238–244
- Models, 238–239, 244–246
 - in chemistry and molecular biology, 246–248
 - simulations versus, 244–246
 - successful, 240–241
- Molecular biology, models in, 246–248
- Molecular bond interactions, 277–280
- Molecular branching, 5, 6–7
- Molecular complexity, 1
- Molecular complexity index, 23, 192
- Molecular dynamics approach, 246–248
- Monocilia, 76
- Monte Carlo calculations, 246–248
- Moore neighborhood, 262, 263
- Morgan algorithm, 17
- Morphogens, 70
- Movement probability, 267
- [M, R] Metabolism–Repair system, 104
- MRDH (mutual repression by dimer and heterodimer) network, 56–58
- Muirhead comparability inequalities, 28
- Multigraph, 194
- Multiports, 130
 - linear, 130–133
- Multistability in cell type diversification, 53–55
 - epigenetic, 55–59

- Mutations, 175
- Mutual repression by dimer and heterodimer (MRDH) network, 56–58
- Nanotube diameter, 43
- Nanotubes
 - complexity of, 36–43
 - helicity of, 38–43
- Neighborhoods, 262–264
- Network complexity
 - measuring, 202–213
 - quantitative measures of, 191–232
- Network thermodynamics, 115–116, 135, 138, 144
 - analog models of, 121
 - electric circuits and, 119–120
 - structure of, 118–133
- Networks, 103, 155, 221
 - assortativeness of, 172
 - biochemical, modeling, 289–297
 - characterizing, 120–121
 - chemical reaction, simulation of, 134
 - dependency, 253
 - formal description of, 126–128
 - gene, 180–187
 - as graphs, 193–201
 - linear resistive, 128–130
 - non-linear, simulation on SPICE, 133–138
 - of protein complexes, 222–226
 - protein interaction, 169–170
 - relational, 138–148
 - topology of, 120, 126
- Newtonian dynamics, 112–114
- Nieuwkoop center, 74
- Nodes, 157, 193
- Non-equilibrium thermodynamics, 141
- Non-linear networks, simulation on SPICE, 133–138
- Non-mechanical behavior, autocatalysis and, 311–316
- Normalized edge complexity, 206
- Objectivity, 100
 - science and, 100–102
- Observables, 242
- OC (overall connectivity), 210–211
- Octane, isomers of, 8, 10, 12–14
- Ohm's law, 123
- Onsager/Prigogine representation, 135
- Onsager reciprocal relations (ORR), 131, 136–138
- Organisms, machines versus, 108–112
- ORR (Onsager reciprocal relations), 131, 136–138
- Oscillatory mechanisms, 84
- Out-adjacency, 196
- Out-center, 200
- Out-component, 200
- Out-degree, 158, 195
- Overall connectivity (OC), 210–211
- Path, 194, 198
- Path graph, 194
- Path matrix, 8
- Pattern, 4
- Pattern formation, embryo, reaction–diffusion mechanisms and, 70–76
- Physical systems, constitutive laws for, 122–123
- Physiology, 105
- Platt index, 207
- Poiseuille's law, 124
- Poisson distribution, 158
- Probabilistic cell movement rules, 266–267
- Propensity, 310
- Protein complexes, networks of, 222–226
- Protein folding maps, small-world structure of, 165
- Protein interaction networks, 169–170
- Protein molecules, 155–156
- Protein structure, 165
 - contact graphs and, 164–169
- Proteins, 164

- Proteome growth, rules of, 176
- Proteome model, 175–180
- Quasi power theorem, 137
- Quest for knowledge, 101
- Random Boolean network (RBN), 181–183
- Random graphs, 158
- RBM (random Boolean network), 181–183
- Reaction–diffusion mechanisms, embryo pattern formation and, 70–76
- Reaction–diffusion systems, 71
- Reductionism, 106, 117–118
 - relational systems theory versus, 107–108
- Relational block diagrams, 103–104
- Relational models of mechanisms, 112
- Relational networks, 138–148
- Relational systems theory, 103
 - in chemistry and biology, 141–144
 - reductionism versus, 107–108
- Relative gravity, 269–271, 272
- Repressors, 55
- Resistive networks, linear, 128–130
- Resistor, 122
- Reversibility, 325
- Riemann zeta function, 186
- Robustness, developmental, evolution of, 83–89
- SC (subgraph count), 207–209
- SC (symmetry corrected), 31
- Schlegel diagrams, 20
- Science
 - mathematics of, 117–118
 - objectivity and, 100–102
- Scientific model, 100–101
- Second cluster coefficient, 197
- Segmentation, 65
 - insect, *see* Insect segmentation
- Self–organization, 254–256
- Self reference, 106
 - Self reference (*cont.*)
 - context dependence and, 102–103
- Sensory physiology, 99
- Shannon entropy, 204
- Shannon theory, 203
- Silicon, 36
- Simple graphs, 194
- Simulations, models versus, 244–246
- Small–world effect, 162–164
- Small–world structure of protein folding maps, 165
- Somitogenesis, 65
 - chemical oscillations and, 65–66
- Somitogenesis oscillator, Lewis–Monk model of, 66–69
- Spanning tree, 194
- Spemann–Mangold organizer, 71–72, 74
- SPICE, non–linear networks simulation on, 133–138
- Star graph, 194
- States of systems, 241–242
- Strongly connected component, 200
- Subgraph, 193
- Subgraph count (SC), 207–209
- Sulfur, 36
- Symmetry
 - fullerenes and, 29–34
 - left–right, axis formation and, 71–72
- Symmetry corrected (SC), 31
- Symmetry factor, 31
- Synchronous cell movement, 264–266
- Synthesis, 109
- System overhead, 322
- Systems, 241
 - states of, 241–242
- TC (topological complexity), 211
- Tellegen’s theorem, 120, 137
- Thermodynamics, 114
 - network, *see* Network thermodynamics
 - non–equilibrium, 141
- Time lag, 90
- TO (topological overlap), 166
- Topological complexity (TC), 211

- Topological indices, 35
- Topological information, 191
- Topological overlap (TO), 166
- Topological reasoning, 116
- Topology
 - of networks, 120, 126
 - vector calculus and, 118
- Total adjacency, 195, 205–206
- Total walk count (TWC), 211–213
- Transcription factors, 55, 180–181
- Trees, 194
- Tube, 226
- Turing instability, 71, 79
- TWC (total walk count), 211–213

- Undirected graph, 194
- Unistors, 132

- Vector calculus, topology and, 118
- Vegetal pole organizer, 73
- Vertex accessibility, 219
- Vertex degree, 195
- Vertex degree distribution, 195, 203
- Vertex distance, 198
- Vertex eccentricity, 198
- Vertices, 157, 193
 - adjacent, 195
- von Neumann neighborhood, 262, 263
 - extended, 264

- Walk, 194
- Walk count, 15
- Walk length, 194
- Water structure, 275–277
- Weighted adjacency matrix, 201
- Weighted graphs, 201
- World, external, human mind and, 99–100

- Zero-flux boundary conditions, 86
- Zeta function, Riemann, 186
- Zigzag nanotubes, 39, 40–42